In the Final Office Action mailed April 3, 2001, the Examiner asserted that the claimed invention lacked utility. In the Amendment submitted July 31, 2001 Applicants addressed this rejection and discussed the association between polyglutamine repeats and neuropsychiatric disorders. Exhibits A-C provided herewith further support the utility of the claimed invention as of the filing date of the above-identified application. In particular, attached Exhibits A-C further demonstrate the association between polyglutamine repeats and neuropsychiatric disorders.

In view of the foregoing, it is submitted that the claims are in condition for allowance. Reconsideration and withdrawal of the rejections is respectfully requested. Should the Examiner have any questions regarding this matter he is invited to telephone the undersigned so that the questions may be resolved.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: Oct. 9, 2001

By: Daniel Hart

Daniel Hart
Registration No. 40,637
Attorney of Record
620 Newport Center Drive
Sixteenth Floor
Newport Beach, CA 92660

S:\DOCS\DOH\DOH-6103.DOC /jc
100801

Glutamine repeats and inherited neurodegenerative diseases: molecular aspects

MF Perutz

Several dominantly inherited, late onset, neurodegenerative diseases are due to expansion of CAG repeats, leading to expansion of glutamine repeats in the affected proteins. These proteins are of very different sizes and, with one exception, show no sequence homology to known proteins or to each other; their functions are unknown. In some, the glutamine repeat starts near the N-terminus, in another near the middle and in another near the C-terminus, but regardless of these differences, no disease has been observed in individuals with fewer than 37 repeats, and absence of disease has never been found in those with more than 41 repeats. Protein constructs with more than 41 repeats are toxic to *E. coli* and to CHO cells in culture, and they elicit ataxia in transgenic mice. These observations argue in favour of a distinct change of structure associated with elongation beyond 37–41 glutamine repeats. The review describes experiments designed to find out what these structures might be and how they could influence the properties of the proteins of which they form part. Poly-L-glutamines form pleated sheets of β-strands held together by hydrogen bonds between their amides. Incorporation of glutamine repeats into a small protein of known structure made it associate irreversibly into oligomers. That association took place during the folding of the protein molecules and led to their becoming firmly interlocked by either strand- or domain-swapping. Thermodynamic considerations suggest that elongation of glutamine repeats beyond a certain length may lead to a phase change from random coils to hydrogen-bonded hairpins. Possible mechanisms of expansion of CAG repeats are discussed in the light of looped DNA model structures.

Address
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Current Opinion in Structural Biology 1996, 6:848–858

Abbreviations

CHO cells	Chinese hamster ovary cells
CI2	chymotrypsin inhibitor-2
DRPLA	dentatorubral-pallidoluysian atrophy
GAPDH	glyceraldehyde phosphate dehydrogenase
GST	glutathione-S-transferase
HD	Huntington's disease
IPTG	isopropylthio-D-galactoside
MJD	Machado-Joseph disease
SBMA	spinal and bulbar muscular dystrophy
SCA	spinocerebellar atrophy
SDS	sodium dodecylsulfate

Introduction

March 1993 saw the discovery of the gene for Huntington's disease (HD), a dominantly inherited, autosomal, severe neurodegenerative disease which generally sets in in middle age. The gene consisted of 67 exons spread over 180 kb of DNA; it contained an open reading frame for a single polypeptide chain of over 3100 amino acids. Exon 1 contained a long repeat of CAGs, coding for glutamines, followed by long repeats of triplets, mostly CCGs, but interrupted by CCAs and CTTs, coding for prolines. The only difference between the normal genes and those found in patients with Huntington's disease consisted in the numbers of CAG repeats: no case of HD has been reported among people with fewer than 37 glutamine repeats, nor has any case with more than 41 glutamine repeats been found free from HD. The severity of the disease increased and the age of onset diminished with the number of CAG repeats [1••]. Earlier, expanded triplet repeats had been recognised as the causes of three other inherited diseases: myotonic dystrophy [2–4], Fragile X syndrome [5,6], and spinal and bulbar muscular atrophy (SBMA), an X-linked syndrome also known as Kennedy disease [7]. The first two are due, respectively, to expansion of CTG and CGG triplets in untranslated regions of the genome, but SBMA was found to arise by an expansion of CAG repeats coding for a repeat of glutamines in the androgen receptor.

Since 1993, several other dominantly inherited, late onset, neurodegenerative diseases have been discovered to be due to expanded glutamine repeats in the affected proteins. Each of them specifically affects different neurones and gives rise to different symptoms. They are Machado-Joseph disease, also known as Spinocerebellar Ataxia Type 3 (MJD or SCA3) [8]; Dentatorubral-pallidoluysian atrophy (DRPLA) [9,10], and Spinocerebellar ataxia type 1 (SCA1) [11]. All three diseases show autosomal dominant inheritance. An antibody specific for expanded glutamine repeats has recently led to the detection of two more autosomal and dominantly inherited cerebellar ataxias, named types 2 and 7 [12]. In a screen of 100128 brain cDNAs for polymorphic CAG and CTG repeats by hybridisation with different primer sequences, Néri *et al.* found three clones of highly polymorphic repeats which could play a part in other neurodegenerative diseases or multifactorial neuropsychiatric disorders [13]. One strange feature common to all these diseases is 'anticipation' by which is meant earlier onset and increased severity in successive generations. This has now been explained by the expansion of the CAG repeats on the mutant chromosomes, especially in male transmission.

These remarkable discoveries posed a great challenge to biomedical research. What is the molecular mechanism of CAG expansion? Can it be prevented? What is the structure and function of the normal gene product?

How does expansion affect them? Why is it toxic to specific neurons in the central nervous system? Can the toxic effects of expanded glutamine repeats be prevented or at least alleviated?

Since 1993 a large literature on these diseases has grown up. In this review I have tried to assemble some of the information that has emerged and to give an account of my colleagues' and my own approach to the problem from the protein-structural point of view.

Huntington's disease

Several observations show that HD is due to a gain rather than a loss of function. For example, patients with a chromosomal translocation which inactivates one of their HD genes exhibited no neurological symptoms, nor do mice in which one of these genes has been knocked out.

The function of the HD protein, now called huntingtin, is unknown, and so is the function of its glutamine repeat. That repeat may not even have a function, because it is reduced to seven glutamines in mice [14] and to only four in the puffer fish (*Fugu rubripes*) [15•]. Yet huntingtin itself must have a vital function, because 'knock-out' mice which lack the HD gene die as 8.5 to 10.5 day-old embryos [16,17]. This function must be preserved even on expansion of the CAG repeat, because patients homozygous for HD grow up normally and exhibit symptoms similar to those of heterozygotes. The gene is expressed in varying degrees in all tissues. Immunostaining of rat brain tissue showed much of the stain to be concentrated in presynaptic terminals. After subcellular fractionation in sucrose density gradients, huntingtin was enriched in a synaptosomal fraction. Its immuno-reactivity dropped with washing with 0.2 M or 1 M NaCl, suggestive of loose association with membranes. Another study of normal and HD-affected human brains also found huntingtin to be membrane-bound, the strongest signal again coming from the synaptosomal fraction [18–22].

Several studies have shown huntingtin to be present only in the cytoplasm, but a study of mouse embryonic and human fibroblasts and mouse neuroblastoma cells showed that an antibody directed against the C-terminal domain of huntingtin stained both nuclei and cytoplasm, but not nucleoli. Cytoplasmic and nuclear protein fractions separated by SDS-PAGE and detected on Western blots also showed huntingtin in both fractions. The authors attribute the failure of others to find huntingtin in the nucleus to differences in antibodies and experimental procedures [23]. The question of its location is important, because most proteins with glutamine repeats are transcription factors, and a nuclear presence opens the possibility that huntingtin might act as a transcription factor-associated protein, even if it lacks an obvious nucleic acid binding site.

The incidence of HD varies widely among ethnic groups, being lowest among South African Blacks (1 in 10^5) and Japanese (0.11–0.45 in 10^5), and 4 in 10^5 on average among Europeans, though varying widely in different countries and districts within countries [24]. Rubinsztein *et al.* [25] have shown that the distribution of CAG repeats in normal populations is not random, but skewed towards lengths above the most frequent length; human CAG repeats are longer than those of primates (7–12 repeats). There is no evidence that longer repeats in humans carry any selective advantage. Rubinsztein *et al.* therefore tried to reproduce the skewed human distribution by a computer model based on random mutations and neutral genetic drift, with a mutational bias towards longer repeats. Their model predicts a mutational bias towards longer alleles over all repeat lengths and an increase of the mutation rate with length, consistent with the finding that new HD mutations originate on chromosomes with high normal repeat numbers. These findings led the authors to the prediction that, over an evolutionary time scale, the lengths of CAG repeats and hence the incidence of HD in human populations is likely to increase.

Machado-Joseph disease (MJD or SCA3)

Machado-Joseph disease is one of the more common inherited spinocerebellar diseases. It is due to expansion of the glutamine repeat in a protein of only about 359 amino acid residues [8]. The repeat lies near the carboxyl end and varies from 13 to 36 glutamines in normal individuals compared to 61 to 84 in affected patients. The remainder of the sequence shows no obvious homology to known proteins and contains no nucleic acid binding site or transmembrane helices [26]. Introduction into CHO cells of the expanded repeats linked to the C-terminal residues alone gradually killed the cells, while introduction of the full-length MJD protein with either normal or expanded repeats, or of the normal-length repeat with the C-terminal residues alone had no effect. On introduction of the same constructs into transgenic mice they were expressed only in Purkinje cells. The expanded repeat with the C-terminal residues alone gave rise to ataxia, gait disturbances and inability to rear, while the constructs with normal repeats had no effect [27].

Western blots of the expressed proteins showed signals at the expected molecular weights for all except the one with the expanded repeat attached to the C-terminal residues. This had higher molecular weights which persisted even after boiling in 2% Na-dodecylsulfate. The authors attribute these to covalent cross-linking of the peptides, but Kelvin Stott and I have found oligo-L-glutamines to form precipitates insoluble in this detergent without any covalent cross-linking (K Stott, M Perutz, unpublished data). The hydrogen bonds between the amide groups alone seem to be sufficient to make these insoluble.

Dentatorubral-pallidoluysian atrophy (DRPLA)

The normal DRPLA gene encodes a protein of about 1184 amino acid residues, with a repeat of 7 to 23 glutamines in normal individuals and expansion from 49 to 75 in affected patients [10]. The repeat begins at residue 484, preceded by five histidines. The protein also contains the polar zipper sequence REREKEARERDLRDR (using amino acid single-letter code), plus several serine repeats and segments rich in serines and prolines, but unfortunately we lack molecular structures to tell us what these signify. RD repeats figure prominently in spliceosomes, which makes the authors suggest that the DRPLA protein may bind RNA. In the human brain the protein is found in the membrane-rich fraction of the neuronal cytoplasm, the same that also contains huntingtin. Rat DRPLA protein has a repeat of only five consecutive glutamines, followed by PQPQPQPQQ [28].

Spinal and bulbar muscular atrophy (SBMA) or Kennedy's disease

This X-linked motor neurone disease is due to expansion of the glutamine repeat in the androgen receptor, a transcription factor of about 918 residues. The repeats begin at residue 59, while the DNA binding site is in the C-terminal domain. The largest normal allele has 36 glutamines and the smallest diseased one 40 [29]. Males who have point mutations in the androgen receptor gene exhibit feminisation, but they do not get motor neurone disease. This is convincing proof that Kennedy disease is due to gain rather than loss of function. On the other hand, males with Kennedy's disease also exhibit mild feminisation.

Kazemi-Esfarjani, Trifior and Pinsky have found the cause of this feminisation [30]. They constructed an androgen receptor-regulated reporter gene and measured its transactivation by an androgen analogue as a function of the length of the CAG repeat in a human androgen receptor cDNA expression vector. They found transactivation to be maximal when all CAGs had been deleted and to decrease steadily with the number of CAG repeats. They suggest that the glutamine repeats allow another, inhibiting protein to bind to the receptor which modulates its activity, but I wonder whether the inhibition is due simply to increasing dimerisation or oligomerisation of the receptor itself.

Spinocerebellar ataxia type 1 (SCA1)

SCA1 is caused by expansion of a glutamine repeat in a protein of about 810 residues. The repeat starts at residue 197. The normal number of glutamines ranges from 19 to 36 and the expanded one from 43 to 81 [31]. In 123 of 126 unexpanded alleles of the SCA1 gene, the CAG repeat was interrupted by at least one CAA triplet, while all expanded alleles from seven different pedigrees had uninterrupted CAG repeats [32]. The largest uninterrupted normal repeat contained 21 CAGs. While this was stable, 70% of transmissions of longer uninterrupted repeats were accompanied by changes in length. Similarly in fragile X-syndrome, which is due to expansion of non-coding CCG repeats, the repeats in normal individuals are interrupted by a single base. The amino acid sequence of the SCA protein shows no homology with known proteins. As in HD, both the normal and the expanded gene are expressed in affected individuals. In non-neuronal cells the protein is found only in the cytoplasm. In Purkinje cells of the cerebellum it is found in both nuclei and cytoplasm, while in other neuronal cells it is found only in the nucleus [33]. The corresponding protein in mice and rats contains only two glutamines. Transgenic mice expressing the human SCA1 gene with 30 CAG repeats had normal Purkinje cells and normal behaviour, while those expressing the gene with 82 repeats developed ataxia and degeneration of Purkinje cells [34].

The amino acid sequences of the MJD, DRPLA and SCA1 proteins show no indication of either nucleic acid binding sites or transmembrane helices.

Table 1 shows that the proteins containing the glutamine repeats are of very different sizes; in two of them the repeats are near the N-termini, in two nearer the middle and in the fifth the repeat lies near the C-terminus. The most striking common feature is that the number of repeats in normal individuals does not exceed 41. On the other hand, several cases of HD with only 37 repeats have been found. They have been attributed to 'greater penetrance' of the gene, but this is no more than a learned name for the enigmatic overlap between normal and pathological lengths of the glutamine repeats [35].

Molecular mechanism of CAG expansion

Deletion or addition of CAG repeats could happen during replication of DNA by either frameshifts (slippage) or recombination. Streisinger and Owen first predicted that frameshift mutations are likely to occur at repetitive sequences, additions by looping-out of the daughter strand and deletions by looping out of the parent strand (Fig. 1) [36]. Their prediction was confirmed by the nucleotide sequences at mutational hot-spots in the lac I gene of *E. coli;* they occurred at tandemly repeated sequences such as CTGGCTGGCTGG, which generated additions and deletions of the repeated tetranucleotide at a rate high compared to other mutational sites ($0.5–2 \times 10^6$) [37]. Replication of double-stranded CAG/CTG triplet repeats *in vitro* was found to be interrupted by pauses which became more marked the longer the repeats [38]. The location of the pauses was determined by the distance between the site of hybridization with the primer and the start of the triplet repeat. They were abolished by heating at 70°C, which is known to melt hairpins and other secondary structure. The pauses may be due to loops or hairpins of the kind predicted by Streisinger and Owen. They may block the DNA polymerase's progress, forcing it to idle and to insert additional triplets while doing so.

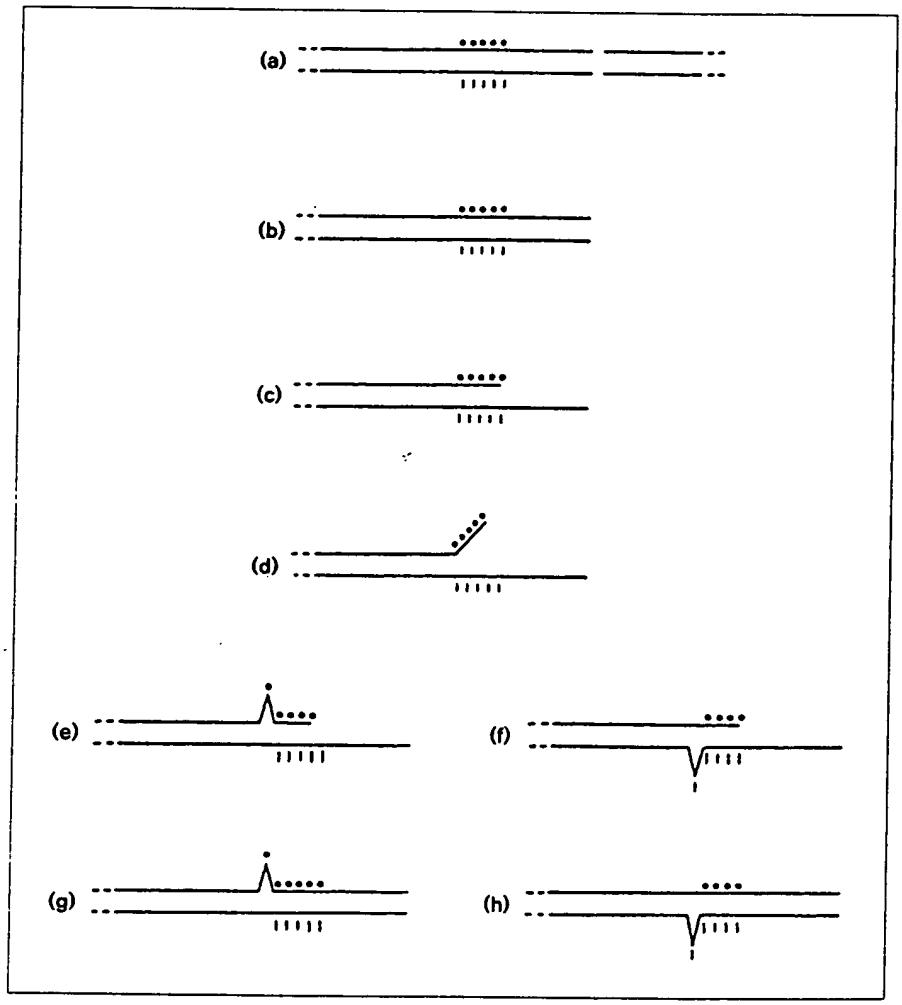
Table 1

Characteristics of glutamine repeats in proteins affected by inherited neurogenerative diseases.

Disease	Approximate number of amino acid residues in normal protein	Start of glutamine repeat at residue number	Longest normal repeat	Shortest repeat in disease
HD	3144	18	41	37•
SMBA	918	59	36	40
SCA1	810	197	39	40
DRPLA	1184	484	39	53
MJD	359	297	36	62

•The CAG repeat of the huntingtin gene is followed by the triplets CAA and CAG. Therefore, the number of glutamines exceeds by two the number of uninterrupted CAG repeats.

Figure 1



Intermediates in the formation of frameshift mutations in double helical DNA. **(a)** Concatameric T4 DNA molecule during intracellular growth, after endonucleolytic cleavage. •s and vertical lines represent a sequence of reiterated base pairs. **(b)** Formation of an end. **(c)** Single-stranded exonuclease digestion. **(d)** Melting of a short terminal stretch. **(e,f)** Annealing in a misaligned position. **(g)** Precursor to an addition mutation. **(h)** Precursor to a deletion mutation. In each diagram the bottom line represents the template and the top line the daughter strand. Note that if the extrahelical base occurs in the daughter strand (e) an addition will be generated, while the bulge defect in the parent strand produces a deletion. Reproduced with permission from [36].

So far as I am aware, no-one has yet demonstrated looping-out of single-stranded hairpins from double-stranded CAG/CTG *in vitro*, but two groups have studied the conformations of single-stranded triplet repeats [39,40]. $(CTG)_{15-25}$ and $(CAG)_{15-25}$ do indeed form stable hairpins in solution, and they are more stable than hairpins formed by other triplet repeats. Interruptions of the regular repeats reduce the stability of the hairpins. This is interesting, seeing that interruption of the CAG or CCG repeats by a single base seems to prevent their fatal expansion in humans.

A little more about the loops can be learnt from Zhu, Chou and Reid's [41] detailed NMR structure of the DNA decamer CAATGCAAATG. It is centred on the triplet GCA, which is related to repeated CAG triplets by a simple frameshift. This decamer forms a hairpin joined at its apex by the single cytidine and held together by interactions between a sheared GA pair (Fig. 2). The authors also studied the structure of the three other GNA triplets inserted into decamers with the same flanking sequences. CGA formed stable antiparallel duplexes while GAA and GTA formed equilibrium mixtures of duplexes

and hairpins; GCA was the only triplet which formed loops made of a single nucleotide (C) held together by a GA pair. The heptadecamer $A(GCA)_5G$ migrated in solution purely as a hairpin. NMR spectra suggested that the hairpins were held together by either G·C or anti-anti G·A pairs. Short $(CTG)_n$ repeats of the strand complementary to $(CAG)_n$ also formed hairpins. The authors suggest that looping-out of both DNA strands could lead to Holliday junctions which play a part in meiotic gene transposition and recombination (Fig. 3).

In hairpins formed by single-stranded CAG or CTG triplet repeats, every third base pair is mismatched, so that they would always be less stable than CAG/CTG double helices in which all base pairs are linked by complementary hydrogen bonds. It seems likely, therefore, that hairpins

Figure 2

Stereo view of the refined CAATGCAATG hairpin structure. The stem consists of four basepairs with a sheared GA pair at each end. Note that 6C stacks nicely on 5G of the 5G7A pair with its sugar stacked on the 7A base. At the open end of the stem, the 5′-overhang 1C stacks similarly on the sheared 10G2A pair which has greater propeller-twist (perhaps due to end effects). Reproduced with permission from [41].

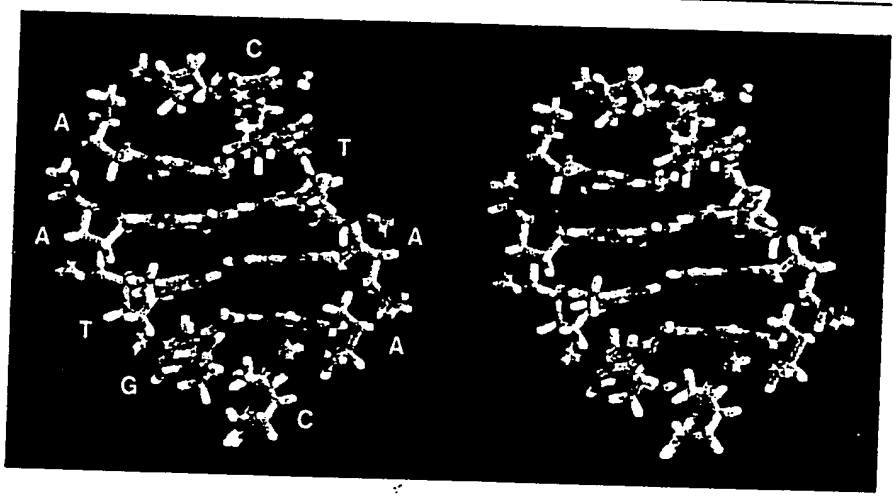
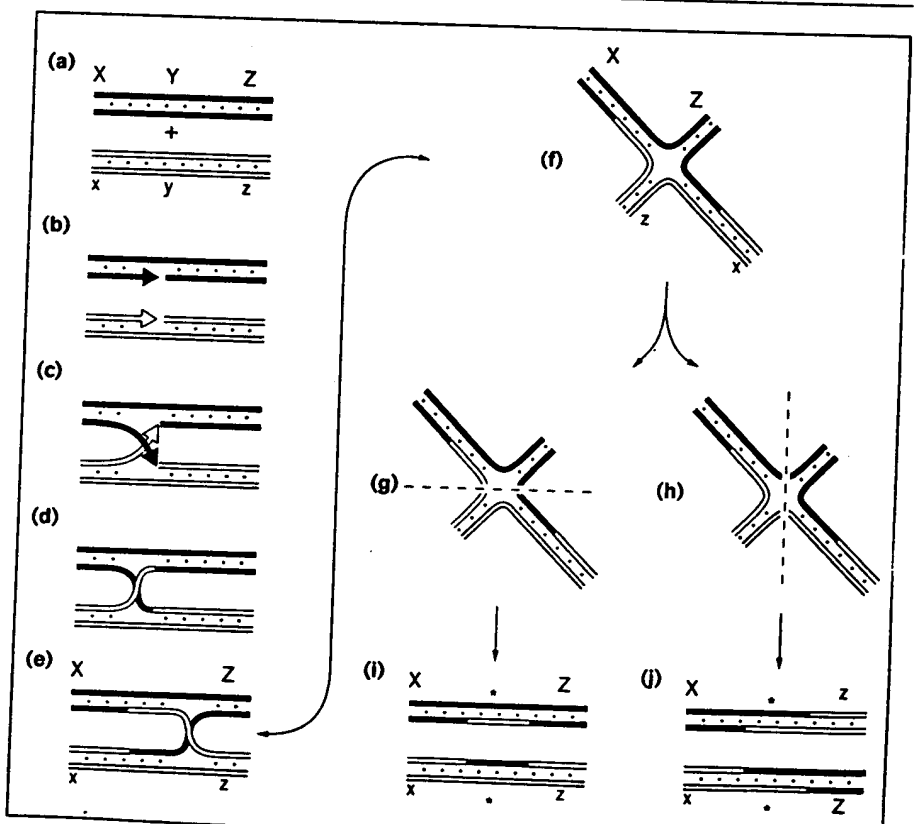


Figure 3

A model of homologous recombination proposed by Robin Holliday [60]. One parental duplex is shown in black and dark grey, and the other in white and pale grey. Either outer strand can base-pair with either inner strand, and vice versa. X, Y, and Z denote three genes; x, y, and z are alleles. The steps are **(a)** alignment, **(b)** cleavage of one strand of each duplex, **(c)** invasion, **(d)** sealing, and **(e)** branch migration. Part **(f)** is an alternative representation of part (e). The lower duplex of (e) is rotated 180 degrees about the vertical axis to give (f). Note that (f) can be cleaved along the **(g)** horizontal or the **(h)** vertical axis, which corresponds to the cleavage of the invading or the non-invading strands, respectively. Rejoining of strands gives two different sets of recombinants **(i)** and **(j)**. Regions containing one strand from each parental duplex, called heteroduplexes, are labeled with an asterisk. After H Potter and D Dressler [61]. Reproduced with permission from [62].



arise only when the double helix is unwound before replication, or during recombination.

Toxicity of long polyglutamine repeats in *E. coli* and mice

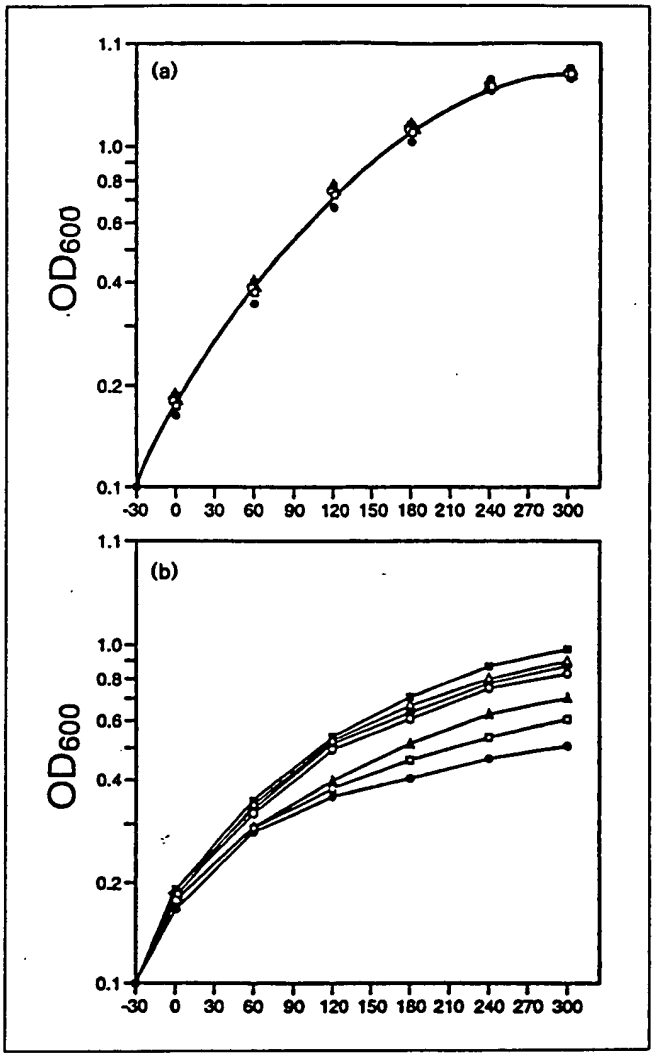
Odoneral *et al.* [42] have constructed fusion proteins of glutathione-S transferase (GST) with glutamine repeats of various lengths whose expression in *E. coli* could be induced by isopropylthio-D-galactoside (IPTG). They found that addition of IPTG to their cultures increased the expression of GST-fusion proteins with 10 to 35 glutamines, but decreased expression of GST-fusion proteins with 59 to 81 glutamines. Suspecting inhibition of cell growth, the authors compared the growth rates of *E. coli* containing constructs without the polyglutamine domain with those containing polyglutamine domains of various length. Before induction with IPTG all these cultures grew at the same rates, but after induction constructs with more than 35 glutamines showed reduced growth rates (Fig. 4). On the other hand, GST-fusion proteins with the equivalent poly-L-alanines had no effect.

Mangiarini *et al.* [43] have bred transgenic mice expressing exon 1 of the human HD gene with different lengths of CAG repeats. This exon coded for the 17 mixed amino-terminal residues, for the glutamine repeat, and finally the two long proline repeats plus the following twelve mixed residues. The authors constructed transgenic lines carrying expansions of 116–156 repeats. In the three lines in which the transgene was expressed, its average expression was 75%, 31% and 77% of the endogenous HD gene. These mice developed a progressive neurological phenotype. In one line, the symptoms developed at 9–11 weeks, and the mice died at between 10 and 13 weeks. Their symptoms included a resting tremor, involuntary and sudden movements resembling chorea, and epileptic seizures induced by handling. The only detectable neuropathology was an average reduction in brain size of 19%, due to a reduction of all brain structures; this resembled early changes associated with HD in humans. It might be argued that the long proline rather than the glutamine repeat produced the symptoms, but this seems unlikely, because that repeat is polymorphic in the mouse, and one allele has a repeat almost as long as the human one. Control mice transfected with an exon 1 containing a repeat of only 18 CAGs are five months old at the time of writing and show no symptoms.

Interaction of huntingtin with other proteins

Li *et al.* [44] searched for proteins with which the rat and human HD proteins associate. They screened a rat brain DNA library by ligating a cDNA encoding the first 230 amino acids of the HD protein to a Gal-4 DNA binding domain plasmid vector of the yeast two-hybrid system and found one positive clone among 100,000 colonies. It had an insert of 600 base pairs encoding part of a new hydrophilic protein which associated with the human HD protein. Expansion of the glutamine repeat of the HD protein

Figure 4



Impairment of growth of *E. coli* expressing GST-polyglutamine fusion proteins. *E. coli* transformed with clones Q(-) (▼); Q10 (■); Q19 (△); Q35 (○); Q59 (▲); Q62 (□); or Q81 (●) were incubated in (a) the absence or (b) the presence of 0.1 mM IPTG or measure of bacterial concentration in the culture medium. Reproduced with permission from [42].

enhanced the association. The full protein of more than 600 amino acid residues contains a repeat of glutamates which would associate with glutamine repeats even more strongly than glutamine repeats would associate with themselves, because hydrogen bonds between amides and carboxylates are stronger than hydrogen bonds between amides. The protein is enriched also in human brain and appears low or absent from other tissues.

Burke *et al.* [45] incubated normal human brain homogenate with polyglutamine-20 and polyglutamine-60 immobilized on resin beads. A protein that was retained only on the polyglutamine-60 beads was identified as the ubiquitous enzyme glyceraldehyde phosphate dehydrogenase (GAPDH). In the converse experiment GAPDH immobilized on agarose beads through its α-amino group was

used to extract proteins from brain extract. It extracted the DRPLA protein. An antibody was raised against the amino-terminal domain of huntingtin which includes the glutamine repeat. This antibody bound several fragments of huntingtin which also bound to GAPDH. The authors suggest that binding of GAPDH by long glutamine repeats in huntingtin or the DRPLA protein may damage neurons by interfering with metabolic pathways requiring that enzyme, but they admit that such a mechanism fails to explain the destruction of only specific cells by proteins with polyglutamine repeats. Koshy *et al.* [46] found similar interactions of GAPDH with the SCA1 protein and the androgen receptor in the yeast two hybrid system and also *in vitro*.

Goldberg *et al.* [47] translated truncated cDNAs of the HD protein which included the CAG repeat and studied its cleavage by a cysteine proteinase active in apoptotic cell death. They found that this enzyme does cleave huntingtin, both *in vitro* and in cultured mammalian cell lines, and that the cleavage rate rises with the length of the glutamine repeat. On the other hand, the increase in the percentage cleaved after three hours' incubation in going from a repeat of 15 glutamines to one of 91 was only 1.3 fold. It may therefore be premature to suggest, as the authors have done, that HD could be due to accelerated apoptotic cleavage of huntingtin in the affected neurons.

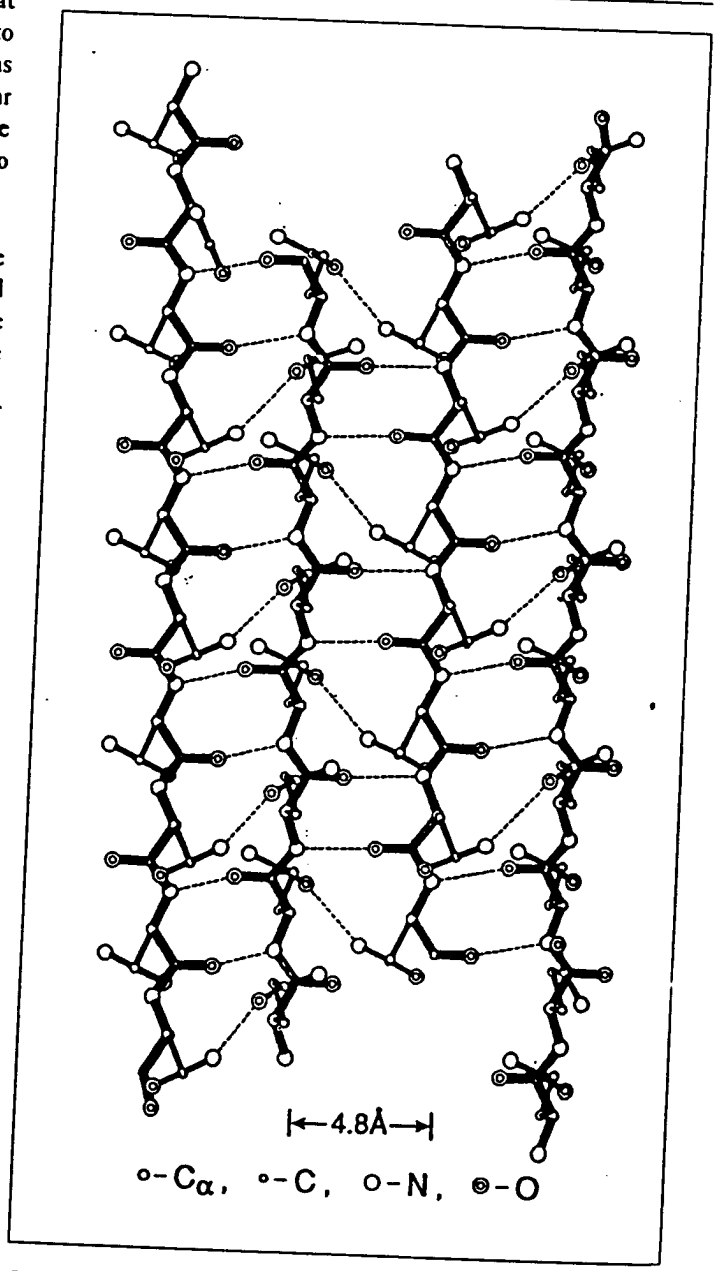
Using the yeast two hybrid system, huntingtin has been found to interact with human ubiquitin-conjugating enzyme which attaches the C-terminal glycine of ubiquitin to specific lysines of its substrate protein [48,49]. Ubiquination is the first step in intracellular breakdown of proteins. In neurons, it might either precede or follow attack by the apoptotic cysteine proteinase, but again the relevance of this reaction to HD is still unclear.

Bao *et al.* [50] have found that huntingtin is retained on calmodulin Sepharose and that such Sepharose retains more mutant than normal protein from brain extracts of HD patients. The physiological relevance of this observation is difficult to judge, because the preparation of huntingtin included purification on an SDS-PAGE gradient which would have denatured the protein.

Structure and function of glutamine repeats

By a strange accident, my attention was drawn to glutamine repeats before the publication of the gene for Huntington's disease. Such repeats have been found, for instance, in some homeodomain proteins of *Drosophila*. A survey of the Swiss Prot Data Bank showed that 33 out of 40 proteins with 20 or more glutamines in a row are transcription factors, many of them in *Drosophila*, involved mostly in the developmental regulation especially of the nervous system [51,52]. Wondering what the structure of glutamine repeats might be, I built an atomic model which showed that β-strands of poly-L-glutamine could be linked together by hydrogen bonds between both their main chain and side chain amides (Fig. 5). In other words, they acted as polar zippers, which made me wonder if they attached β-strands of proteins to each other, while leucine zippers had evolved to make α-helices stick together [53].

Figure 5



Computer-generated structure of two paired antiparallel β-strands of poly(L-glutamine) linked together by hydrogen bonds between the main-chain and side-chain amides.

When I read the astonishing paper in *Cell* on the gene for Huntington's disease, it occurred to me that the polar zipper action of glutamine repeats might furnish a possible clue to the molecular mechanism of the disease, but in view of the medical importance of the problem it seemed essential to test that idea experimentally [illegible]

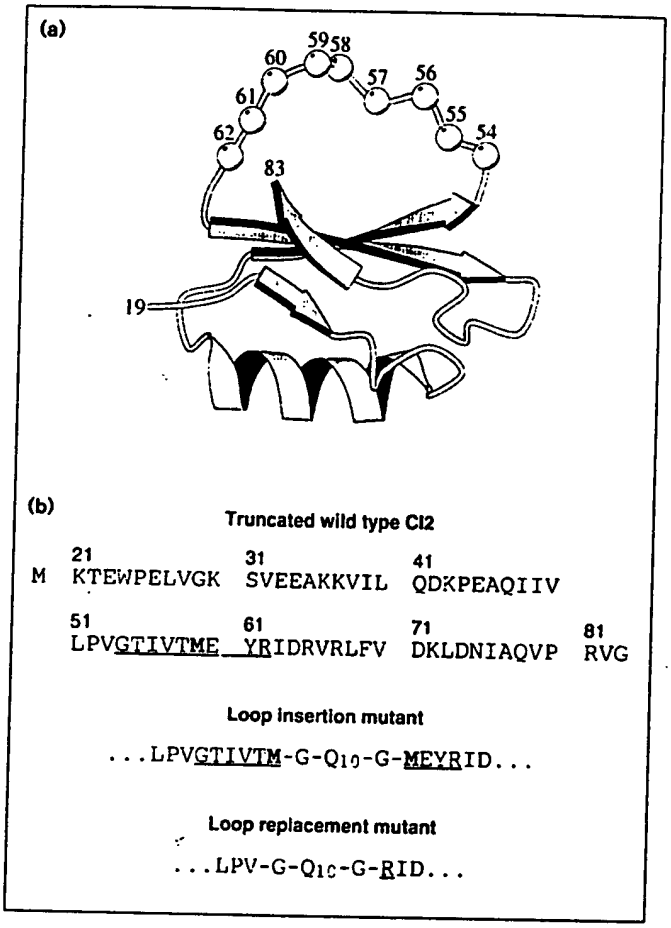
a synthetic polypeptide and then a protein. Poly-L-glutamine alone is insoluble in water. To render it soluble, my colleague Tony Johnson synthesized the polypeptide Asp_2–Gln_{15}–Lys_2. This formed large aggregates which circular dichroism, electron-, and X-ray diffraction proved to consist of β-pleated sheets, consistent with my polar zipper idea, even though isolated glutamines in proteins tend to favour α-helix formation [54].

It remained to find out whether incorporation of glutamine repeats into a small monomeric protein of known structure would make it associate into oligomers. On AR Fersht's advice, I chose chymotrypsin inhibitor-2 (CI2) from barley seeds, a small protein with an external loop of nine residues carrying the inhibitory methionine. Kelvin Stott, Jonathan Blackburn and I engineered two constructs of CI2 from which the first 20 disordered residues had been deleted. In one construct, we replaced the loop by the sequence Gly–Gln_{10}–Gly, and in the other we inserted that sequence into the loop (Fig. 6). The two constructs, together with the truncated wild-type protein, were expressed in good yield in *E. coli*. Fractionation of the cell extracts by fast protein liquid chromatography yielded the truncated wild type as a single fraction, while each of the constructs yielded two additional fractions of higher molecular weight. Analytical ultra-centrifugation showed that the wild type fraction contained only monomers, whereas the additional fractions contained in the two constructs represented dimers and trimers, as predicted if the glutamine repeats acted as polar zippers [55].

It seemed important to determine the dissociation constants of the CI2 oligomers, but they proved stable up to 70°C, above which they denatured. Nor did they dissociate in urea at concentrations below 3 M; at higher urea concentrations they denatured cooperatively with their glutamine repeats. Suspecting that this was a consequence of the large number of hydrogen bonds between the row of ten glutamines, Teca Galvao made constructs of the inhibitor with only 8, 6 or 4 glutamines inserted into its loop. Contrary to expectation, reduction of the number of glutamines only reduced the yields of dimers and trimers, but not their stability, which suggested that the glutamines made these protein molecules interlock in some manner while they were folding in the *E. coli* cell. To test this idea, Stott denatured a concentrated solution of the monomers in 4M guanidinium-HCl and then allowed them to refold by dialyzing the solution against water. The result was a mixture of monomers, dimers and trimers, as found in the *E. coli* extract, which confirmed that these oligomers do indeed form during folding of the protein.

Stott now began to wonder whether the formation of oligomers was specific for glutamines, or whether it was a consequence simply of our having enlarged the external loop and would occur also if other amino acid residues were inserted into it. He therefore engineered the trypsin inhibitor with an insert of ten glycines and

Figure 6



(a) Structure of wild-type CI2; the first 18 residues are disordered and are not shown [7]. Loop residues 54–62 are represented by spheres. The Gly–Gln_{10}–Gly peptide was inserted into the loop at Met59 in one mutant, while the second mutant was generated by replacing residues 54–61 of the loop with this peptide. **(b)** Amino acid sequence of the truncated wild-type CI2 and of the two mutants. Original residues of the loop are underlined. Reproduced with permission from [55].

serines instead of glutamines. This construct did form dimers and trimers, but in the reversible denaturation test the ratios of dimers/monomers and trimers/monomers formed were respectively one third and one ninth of those with glutamine repeats. In another test with an insert of alanines and serines engineered by Petra Scamborova these ratios were only slightly lower than those with glutamine repeats, showing that van der Waal's interactions between the methyl group of the alanines and/or hydrogen bonds between the γ-hydroxyls of serine could also stabilize interactions between the CI2s. In each construct association was again irreversible short of denaturing the proteins, showing that the oligomers formed by some manner of interlocking between monomers during folding. Formation of irreversibly linked oligomers may be due either to domain-swapping, as found in several proteins [56], or more likely to strand exchange, as found in α_1-antitrypsin where an external loop of one molecule

inserts itself into the β-sheet of a neighbouring molecule [57]. The answer awaits structural analysis.

As far as the diseases are concerned, the most obvious inference would be that the glutamine repeats cause self-association into large protein aggregates, but immunostaining of the degenerate neurons in post-mortems of HD patients has shown no evidence for these. Alternatively, association with other proteins may poison the cells.

I now come to the most remarkable property of proteins with glutamine repeats, namely the toxicity of repeats with more than 41 consecutive glutamines, not only to certain neuronal cells, but even to *E. coli* and some mammalian cells in culture. This abrupt transition to toxicity suggests that elongation beyond 41 is associated with a phase change, i.e. a transition between two distinct structures. Isolation of a monoclonal antibody which recognises specifically expanded glutamine repeats in huntingtin and in the proteins of SCA1 and 3 also speaks in favour of a phase change, i.e. a structural transition caused by expansion [58••]. Thermodynamic considerations suggest what this transition might be.

A preliminary NMR study of the structure of CI2 with an insert of ten glutamines showed them to have a random structure. In a random coil, the high entropy of the glutamine repeat itself would have a stabilizing effect; on the other hand, its free amides would bind and immobilize water; this would diminish the entropy of the system and therefore its stability. On the other hand, in a rigid hairpin with hydrogen bonds between all the amides, such as shown in Figure 5, the entropy of the polyglutamine repeat itself would be low, but the liberation of water molecules would compensate for this by increasing the entropy of the system of protein plus water. These two effects may balance each other up to a critical length when another effect would enter. This is known as the chelate effect and is a major stabilizing factor of protein and nucleic acid structures.

With each pair of glutamines joined together to form a hairpin, the loss of translational and rotational entropy suffered on joining the next pair of glutamines would be lessened. At a critical length, that loss may become negligibly small and the system would then become strongly stabilised by the gain in entropy due to the liberated water molecules. Hairpins may therefore become dominant when the number of glutamines exceeds 41. These ideas will now have to be tested with suitable protein models. Hairpins may insert themselves into other protein molecules of either the same or a different kind and cause interactions which poison the cell, perhaps by the formation of oligomers during protein folding.

It seems that the formation of oligomers by insertion of the β-strand of one molecule into the β-sheet of another molecule during their partial unfolding and refolding via a 'molten globule' state may cause other diseases. The polymerisation of prions responsible for scrapie, and by implication for bovine spongiform encephalitis and Jakob-Creutzfeld disease, is accompanied by a transition from a mainly α-helical to a β-sheet structure, induced by infectious prions with a β-sheet structure. A fatal, human, familial amyloidosis was found to be caused by the aggregation of soluble, partly α-helical mutant lysozyme molecules into insoluble fibrils with a β-sheet structure ([59] and M Sunde, P Pepys, CM Dobson, personal communication). Our observations may therefore lead us to a better understanding of the molecular pathology also of amyloid diseases.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Huntington's Disease Collaborative Research Group: **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes.** *Cell* 1993, 72:971–983.

This paper reports the discovery of the gene for Huntington's disease, and shows that the only difference between the normal gene and the gene in affected patients consists in an expansion of a long repeat of CAGs, coding for a repeat of glutamines close to the amino terminus of the protein encoded by the gene.

2. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H *et al.*: **Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3'end of a transcript encoding a protein kinase family member.** *Cell* 1992, 8:799–806.

3. Fu Y-H, Pizzuti A, Fenwick RG, King J, Rajnarayan S, Dunne PW, Dubel J *et al.*: **An unstable triplet repeat in a gene related to myotonic muscular dystrophy.** *Science* 1992, 255:1256–1258.

4. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G *et al.*: **Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene.** *Science* 1992, 255:1253–1255.

5. Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A *et al.*: **Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X syndrome.** *Cell* 1991, 65:905–914.

6. Fu Y-H, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S *et al.*: **Variation of the CGG repeat at the Fragile X site results in genetic instability: resolution of the Sherman Paradox.** *Cell* 1991, 67:1047–1058.

7. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fishbeck KH: **Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.** *Nature* 1991, 352:77–79.

8. Kawaguchi Y, Okamoto T, Taniwaki A, Aizawa M, Inoue M, Katayama S *et al.*: **CAG expansions in a novel gene for Machado-Joseph Disease at chromosome 14q32.1.** *Nat Genet* 1994, 8:221–227.

9. Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K *et al.*: **Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA).** *Nat Genet* 1994, 6:9–13.

10. Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M *et al.*: **Dentatorubral and pallidoluysian atrophy: expansion of an unstable CAG trinucleotide on chromosome 12.** *Nat Genet* 1994, 6:14–18.

11. Orr RT, Chung M-Y, Banfi S, Kwiatkowsky TJ, Servadio A, Beaudet AL *et al.*: **Expansion of the unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1.** *Nat Genet* 1993, 4:221–226.

12. Trottier, Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G *et al.*: Polyglutamine expansion as a pathological epitope in Huntington's Disease and four dominant cerebellar ataxias. *Nature* 1995, 378:403–406.

13. Nèri C, Albanèse V, Lebre A-S, Holbert S, Saada C, Bouguelevet L, Meier-Ewart S: Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders. *Hum Mol Genet* 1996, 5:1001–1009.

14. Lin B, Nasir J, Macdonald H, Hutchinson G, Graham RK, Rommens JM, Hayden MR: Sequence of the murine Huntington Disease gene: Evidence for conservation, and polymorphism in a triplet (CCG) repeat alternate splicing. *Hum Mol Genet* 1994, 3:85–92.

15. Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G *et al.*: Comparative sequence analysis of the human and puffer fish Huntington's disease genes. *Nat Genet* 1995, 10:67–71.
This paper shows that the gene which is mutated in Huntington's disease is common to all vertebrates, but that the long CAG repeat is not, suggesting that the long glutamine repeat in the human protein may serve no function.

16. Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM *et al.*: Inactivation of the mouse Huntington's Disease gene homolog Hdh. *Science* 1995, 269:407–410.

17. Zeilin S, Liu J-P, Chapman DL, Papaioannou VE, Efstratiadis A: Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat Genet* 1995, 11:155–163.

18. Sharp AH, Loev SJ, Schilling G, Li S-H, Li X-J, Bao J: Widespread expression of Huntington's Disease gene (IT5) protein product. *Neuron* 1995, 14:1065–1074.

19. Schilling G, Sharp AH, Loev SJ, Wagster MV, Li S-H, Stine OC *et al.*: Expression of the Huntington's disease (IT5) protein product in HD patients. *Hum Mol Genet* 1995, 4:1365–1371.

20. Persichetti F, Ambrose CM, Pei G, McNeil SM, Srinidhi J, Anderson MA, Jenkins B *et al.*: Normal and expanded Huntington's disease gene alleles produce distinguishable proteins due to translation across the CAG repeat. *Mol Med* 1995, 1:374–383.

21. Trottier Y, Devis D, Imbert G, Saudou F, An I, Lutz Y *et al.*: Cellular localisation of the Huntington's disease protein and discrimination of the normal and mutated form. *Nat Genet* 1995, 10:104–110.

22. Wood JD, MacMillan JC, Harper PS, Lowenstein PR, Jones AL: Partial characterisation of murine huntingtin and apparent variations in the subcellular localisation of huntingtin in human, mouse and rat brain. *Hum Mol Genet* 1996, 5:481–487.

23. De Rooij KE, Dorman JC, Smoor MA, Den Dunnen JT, Van Ommen G-JB: Subcellular localisation of the Huntington's disease gene product in cell lines by immunofluorescence and biochemical subcellular fractionation. *Hum Mol Genet* 1996, 5:1093–1099.

24. Harper P: The epidemiology of Huntington's disease. In *Huntington's Disease*. Edited by Harper P. London: WB Saunder Company Ltd; 1991:251–280.

25. Rubinsztein DC, Amos W, Goodburn S, Rajkumar RR, Old J, Bontrop R *et al.*: Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat Genet* 1994, 7:525–530.

26. Rubinsztein DC, Leggo J, Coetze GA, Irvine RA, Buckley M, Ferguson-Smith MA: Sequence variation and size ranges of CAG repeats in Machado-Joseph disease, spinocerebellar ataxia type 1 and androgen receptor. *Hum Mol Genet* 1995, 4:1585–1590.

27. Ikeda H, Yamaguchi M, Sugai S, Aze Y, Narumiya S, Kukizuka A: Expanded polyglutamine in the Machado-Joseph disease protein induces cell death *in vitro* and *in vivo*. *Nat Genet* 1996, 13:196–202.

28. Schmitt I, Epplen JT, Riess O: Predominant neural expression of the gene responsible for dentatorubral-pallidoluysian atrophy (DRPLA) in rats. *Hum Mol Genet* 1995, 4:1619–1624.

29. Tilley WD, Marcelli M, Wilson JD, McPhaul MJ: Characterisation and expression of a cDNA encoding the human androgen receptor. *Proc Natl Acad Sci USA* 1989, 86:327–331.

30. Kazemi-Esfarjani P, Trifiro MA, Pinsky L: Evidence for a repressive function of the long glutamine tract in the human androgen receptor: possible pathogenic relevance for the $(CAG)_n$-expanded neuronopathies. *Hum Mol Genet* 1995, 4:523–527.

31. Banfi S, Servadio A, Chung M-Y, Kwiatkowski TJ Jr, McCall AE, Duvick LA *et al.*: Identification and characterisation of the gene causing type 1 spinocerebellar ataxia. *Nat Genet* 1994, 7:513–520.

32. Chung M-Y, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT: Evidence for a mechanism predisposing to intergenerational CAG instability in spinocerebellar ataxia type 1. *Nat Genet* 1993, 5:254–258.

33. Servadio A, Koshy B, Armstrong D, Antalfi B, Orr HT, Zoghbi HY: Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nat Genet* 1995, 10:94–98.

34. Burright EN, Clark BR, Servadio A, Matilla T, Fedderson RM, Yunis WS: SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* 1995, 82:937–948. [Quoting unpublished data from S Banfi and HY Zoghbi.]

35. Rubinsztein DC, Leggo J, Coles R, Almquist E, Biancalana V, Cassiman J-J: Phenotypic characterisation of individuals with 30–40 CAG repeats in the Huntington's disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. *Am J Hum Genet* 1996, 59:16–22.

36. Streisinger S, Owen J: Mechanism of spontaneous and induced mutation in bacteriophage T4. *Genetics* 1985, 109:633–659.

37. Farabough PJ, Miller JH: Genetic studies of the lac repressor VIII. On the molecular nature of spontaneous hotspots in the lacI gene of *Escherichia coli*. *J Mol Biol* 1978, 126:847–863.

38. Kang S, Ohshimak, Shimuzu M, Amirhaeri A, Wells RD: Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J Biol Chem* 1995, 270:27014–27021.

39. Gacy AM, Goellner J, Jurani N, Macura S, McMurray CT: Trinucleotide repeats that expand in human disease from hairpin structures *in vitro*. *Cell* 1995, 81:533–540.

40. Mitas M, Yu A, Dill J, Kamp TJ, Chambers EJ, Haworth IS: Hairpin properties of single-stranded DNA containing a GC-rich repeat. *Nucleic Acids Res* 1995, 23:1050–1059.

41. Zhu I, Chou S-H, Xu J, Reid BR: Structure of single cytidine loop formed by the DNA triplet GCA. *Nat Struct Biol* 1995, 2:1012–1017.

42. Onoderal O, Roses AD, Tsuji S, Vance JM, Strittmater WJ, Burke JR: Toxicity of expanded polyglutamine domain proteins in *Escherichia coli*. *FEBS Lett* 1996, in press.

43. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C *et al.*: Exon 1 of the Huntington's disease gene containing a highly expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 1996, in press.

44. Li X-J, Li S-H, Sharp AH, Nucifom FC Jr, Schilling G, Lanahan A: A Huntington-associated protein enriched in brain with implications for pathology. *Nature* 1995, 378:398–402.

45. Burke JR, Enghild JJ, Martin ME, Jou Y-S, Myers RM, Roses AD *et al.*: Huntingtin and DRPLA proteins selectively interact with the enzyme GAPDH. *Nat Med* 1996, 2:347–350.

46. Koshi B, Matilla T, Burright EN, Merry DE, Fishbeck KH, Orr HT, Zoghbi HY: Spinocerebellar alaxia type-1 and spinobulbar muscular atrophy gene products interact with glyceraldehyde-3-phosphate dehydrogenase. *Hum Mol Genet* 1996, 5:1311–1318.

47. Goldberg YP, Nicholson DW, Rasper DM, Kalchman MA, Koide HB, Graham RK: Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nat Genet* 1996, 13:442–449.

48. Kalchman MA, Graham RK, Xia G, Koide HB, Hodgson JG, Graham KC *et al.*: Huntingtin is ubiquinated and interacts with a specific ubiquity conjugating enzyme. *J Biol Chem* 1996, in press.

49. Nasir J, Goldberg YP, Haydon MR: Huntington disease: new insights into the relationship between CAG expansion and disease. *Hum Mol Genet* 1996, 5:1431–1435.

50. Bao J, Sharp AH, Wagster MV, Becker M, Schilling G *et al.*: Expansion of polyglutamine repeat in huntingtin leads to

abnormal protein interactions involving calmodulin. *Proc Natl Acad Sci USA* 1996, 93:5037–5042.

51. Gerber HP, Seipel K, Giorgiev O, Hofferer M, Hug M, Rusconi S: Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 1994, 263:808–811.

52. Karlin S, Bueger C: Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc Natl Acad Sci USA* 1996, 93:1560–1565.

53. Perutz MF, Staden R, Moens L, De Baere I: Polar zippers. *Curr Biol* 1993, 3:249–253.

54. Perutz MF, Johnston T, Suzuki M, Finch JT: Glutamine repeats as polar zippers: their possible role in neurodegenerative diseases. *Proc Natl Acad Sci USA* 1994, 91:5355–5358.

55. Stott K, Blackburn JM, Butler PJG, Perutz MF: Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proc Natl Acad Sci USA* 1995, 92:6509–6513.

56. Bennett MJ, Schlunegger MP, Eisenberg D: 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* 1995, 4:2455–2468.

57. Carrell RW, Stein PE: The biostructural pathology of the serpins: critical funxtion of sheet opening mechanism. *Biol Chem Hoppe Seyler* 1996, 377:1–17.

58. •• Trottier Y, Lutz Y, Stevanin G, Imbert G, Derys D, Cancel G: Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* 1995, 378:403–406.

This paper reports the discovery of a monoclonal antibody which recognises specifically expanded glutamine repeats, suggesting that expansion causes a change of structure.

59. Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK: Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* 1993, 362:553–557.

60. Holliday R: A mechanism of gene conversion in fungi. *Genet Res* 1964, 5:282.

61. Potter H, Dressler D. *Cold Spring Harbor Symp Quant Biol* 1979, 43:973.

62. Stryer L: *Biochemistry*, edn 4. New York: WH Freeman & Co:820.

EXHIBIT B

Anticipation in Swedish families with bipolar affective disorder

Per-Olof Nylander, Christer Engström, Jayanti Chotai, Jan Wahlström, Rolf Adolfsson

Abstract

Anticipation describes an inheritance pattern within a pedigree with an increase in disease severity or decrease in age at onset or both in successive generations. The phenomenon of anticipation has recently been shown to be correlated with the expansion of trinucleotide repeat sequences in different disorders. We have studied differences of age at onset and disease severity between two generations in 14 families with unilinear inheritance of bipolar affective disorder (BPAD).

There was a significant difference in age at onset ($p<0.008$), in episodes per year with ($p<0.006$) and without ($p<0.03$) lithium treatment, and in total episodes per year ($p<0.002$) between generations I and II. Furthermore, there was a highly significant correlation ($p<0.001$) in age at onset between generations I and II. No evidence for specific paternal or maternal inheritance was found.

We found evidence of anticipation and could rule out ascertainment bias or some other artefact. Anticipation is thus an inheritance pattern in BPAD which suggests that the expansion of trinucleotide repeat sequences is a possible mode of inheritance in BPAD.

(*J Med Genet* 1994;31:686–689)

Bipolar affective disorder (BPAD) is a severe neuropsychiatric disorder with a prevalence of about 1% in the population. Previous family,[1 2] twin,[3] and adoption studies[4] support the hypothesis that genetic transmission is a major aetiological factor in the disease. Linkage analysis has suggested gene localisation to chromosome 11 for BPAD,[5] but further studies have not been able to confirm these results and the results have been inconsistent.[6-8] The mode of inheritance of BPAD is still unknown but the recent demonstration of anticipation in BPAD opens up new possibilities for studying the genetic transmission of this disorder.[9]

Anticipation describes an inheritance pattern within a pedigree with an increase in disease severity or decrease in age at onset or both in successive generations.[10 11] In early clinical studies anticipation was suggested as an inheritance model of psychiatric disorders.[12-14] Anticipation has been found in myotonic dystrophy,[10 15] fragile X syndrome,[16] Huntington's disease,[17-20] amyloidosis,[21] and BPAD.[9] Anticipation has been shown to be directly associated with the expansion of trinucleotide repeats at the disease locus in fragile X syndrome,[22] myotonic dystrophy,[23 24] and Huntington's disease.[17-20]

In this study we have examined 13 two generation unilinear BPAD family pairs and one BPAD family affected in three successive generations, to look for evidence of anticipation. Our results are in agreement with anticipation as an inheritance pattern in pedigrees of BPAD, resulting in an increase in disease severity and a decrease in age at onset in successive generations.

Material and methods

BPAD patients were recruited from the lithium dispensaries at the Psychiatric Clinics of Umeå and Härnösand Hospitals in northern Sweden. Lithium therapy is used as a prophylaxis in BPAD patients, aiming to reduce the frequency and severity of depressive/manic episodes. Its feasibility depends on compliance from the patient. From the register we found that about 25% of these patients fulfilled the DSM-IIIR criteria of BPAD (types I and II). The patient files were checked by OPCRIT,[25] which further confirmed the diagnosis.

Information about the occurrence of psychiatric morbidity in the relatives of these patients was obtained from medical records. In the present study we included only pedigrees with unilinear inheritance of BPAD. Cases of BPAD with other diagnoses in their pedigrees such as schizophrenia, schizoaffective syndrome, unipolar major depression, or unipolar mania were excluded from the study. Altogether we found 14 pedigrees with unilinear inheritance of BPAD, one of these showing three generation inheritance. In five of these pedigrees, one parent and two children each had BPAD.

Anticipation was measured as (1) age at onset and (2) frequency of episodes with and without lithium treatment. The age at onset was defined by the first episode of either depression or mania. Episode frequency was used as a measure of disease severity and its overall value was calculated by the total number of episodes of depression and mania divided by the total number of years from age at onset to death or to the time of investigation. For each person, information was also obtained regarding the episode frequency and the length of time with or without lithium treatment, respectively. This enabled separate estimates of episode frequency either during lithium treatment or when there was no lithium treatment.

Department of Psychiatry, University of Umeå, S-901 85 Umeå, Sweden
P-O Nylander
C Engström
J Chotai
R Adolfsson

Department of Clinical Genetics, University of Umeå, Sweden
P-O Nylander

Department of Clinical Genetics, Gothenborg, Sweden
J Wahlström

Correspondence to Dr Nylander.

Received 28 September 1993
Revised version accepted for publication 19 April 1994

Family	Generation	Sex	Year of birth	Age at onset (y)	Episodes/y no lithium	Episodes/year lithium	Total episodes/year
1	I	F	1932	30	0·1	–	0·1
	II	M	1952	32	1·0	1·0	1·0
	II	F	1953	21	1·3	0·5	0·8
2	I	F	1928	30	0·3	0·0	0·2
	II	F	1955	18	0·6	2·0	0·9
3	I	M	1940	25	1·5	0·5	0·7
	II	M	1974	18	1·0	–	1·0
4	I	F	1917	35	0·1	0·1	0·1
	II	M	1941	31	1·0	0·4	0·4
5	I	M	1911	52	0·4	–	0·4
	II	M	1946	37	1·3	0·7	0·9
6	I	F	1927	49	0·6	–	0·6
	II	F	1948	36	?	?	2·0
7	I	F	1926	34	0·6	0·1	0·4
	II	F	1950	14	0·6	0·2	0·4
8	I	F	1927	38	1·0	0·0	0·1
	II	F	1952	23	3·0	0·1	0·2
	II	M	1954	21	0·5	1·0	0·6
9	I	F	1901	33	0·2	–	0·2
	II	F	1935	29	0·3	0·6	0·5
	III	M	1958	23	1·0	1·4	1·3
10	I	M	1921	34	0·4	–	0·4
	II	M	1950	22	0·2	0·0	0·2
	II	M	1956	22	0·4	–	0·4
11	I	F	1892	50	0·4	–	0·4
	II	F	1913	44	0·5	0·8	0·7
12	I	M	1921	47	0·5	0·0	0·2
	II	F	1948	36	?	?	0·5
	II	M	1953	22	?	?	0·4
13	I	M	1936	20	0·4	0·0	0·3
	II	M	1973	15	2·5	–	2·5
14	I	M	1915	21	0·7	0·0	0·4
	II	M	1946	16	0·6	–	0·6
	II	F	1952	20	0·4	–	0·4

F = female, M = male, ? = unknown, – = no treatment with lithium

Table 2 Episodes per year in generations I and II with and without lithium treatment

Treatment	Generation	No	Episodes per year		p
			Range	Mean (SD)	
No lithium	I	15	0·1–1·5	0·50 (0·36)	
	II	17	0·2–3·0	0·95 (0·76)	<0·03
With lithium	I	9	0·0–0·6	0·14 (0·23)	
	II	12	0·0–2·0	0·72 (0·57)	<0·006
Total	I	15	0·1–0·7	0·33 (0·18)	
	II	20	0·2–2·5	0·78 (0·58)	<0·002

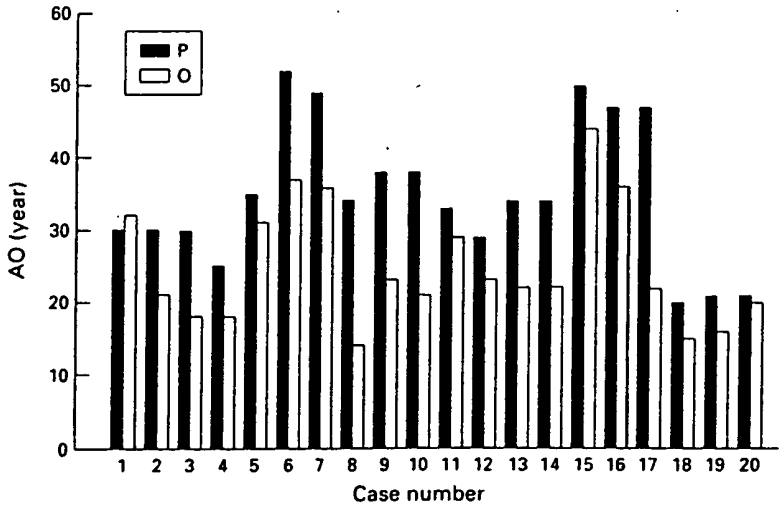


Figure 1 The differences in age at onset (AO) between the 20 family pairs with BPAD. P = parent, O = offspring.

STATISTICAL ANALYSIS

The three generation family was broken up into two generation pairs, so the person in the middle generation appeared once as a parent and once as a child. For comparing generation I and generation II, the Wilcoxon rank sum test for independent samples was used to test for statistical significance. To investigate paternal or maternal inheritance and for the correlation analyses, however, pairwise comparisons were made within parent-offspring pairs. Each of the five families with two BPAD children were treated as two parent-offspring pairs. Statistical significance was tested by the Wilcoxon rank sum test for pair observations. Correlation analyses were performed by Spearman correlation coefficients. Two tailed p values were used for all tests.

Results

The family data including sex, generation, year of birth, age at onset, total episodes per year, and episodes per year with and without lithium treatment are shown in table 1.

In generation I (n = 15) the range for birth year was 1892–1940 and for age at onset 20–52 years. The corresponding figures for generation II (n = 20) were 1913–1974 and 14–44 years, respectively. The mean age at onset was 35·1 (SD 10·3) in generation I and 25·0 (SD 8·4) in generation II. There was a significant difference ($p<0·008$) in age at onset (10·1 years) between generations I and II.

The number of episodes per year in generations I and II, with and without lithium treatment, can be seen in table 2. A significantly higher frequency ($p<0·002$) of total episodes per year was found in generation II (0·78, SD 0·58) compared with generation I (0·33, SD 0·18). There were also significant differences in episodes per year with ($p<0·006$) and without ($p<0·03$) lithium treatment between generations I and II, with the highest frequencies in generation II.

The distribution of the age at onset (AO) for the 20 parent-offspring pairs can be seen in fig 1. Anticipation (age at onset in parent minus

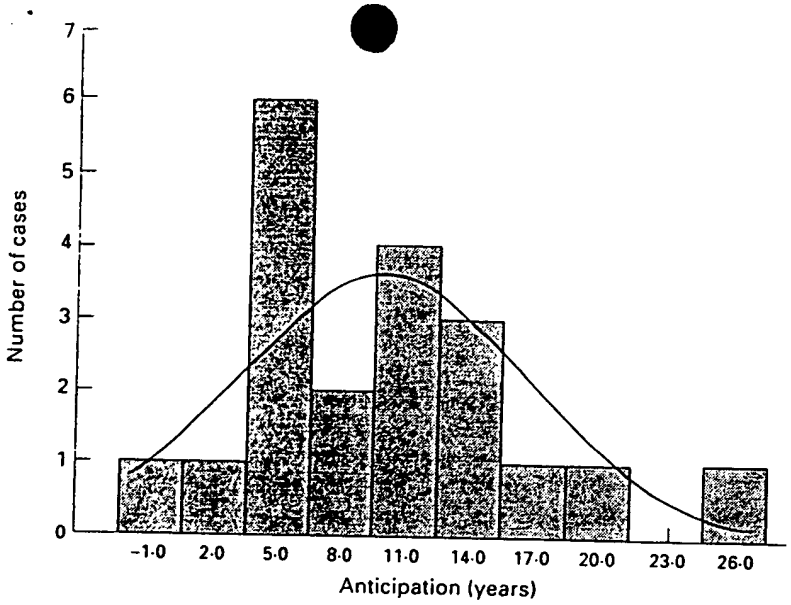


Figure 2 Distribution of anticipation (age at onset in parent minus age at onset in offspring) of 20 cases with BPAD. A normal distribution curve is plotted.

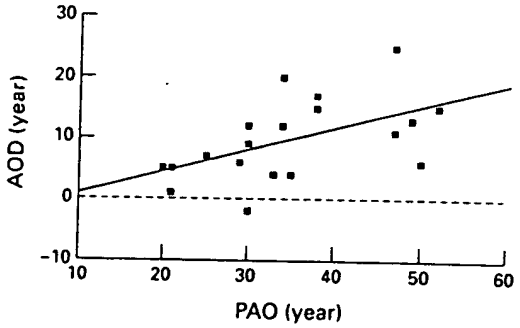


Figure 3 Regression analyses between anticipation (age at onset in parent minus age at onset in offspring, AOD) and parental age at onset (PAO).

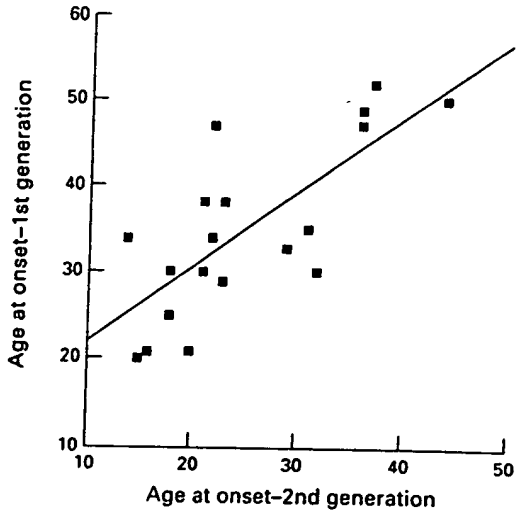


Figure 4 Correlation analyses for age at onset (year) between generations I and II.

Table 3 Paternal and maternal inheritance of bipolar affective disorder: age at onset

Generation pair	No	Age at onset		Diff	p
		Range (y)	Mean (SD)		
Father–offspring	9				
Father		20–52	33·4 (12·6)		
Offspring		15–37	23·1 (8·0)	10·3 (7·0)	<0·008
Mother–offspring	11				
Mother		30–50	36·0 (7·4)		
Offspring		14–44	26·5 (8·8)	9·5 (6·6)	<0·005
Father–son	7				
Father		20–52	33·3 (12·5)		
Son		15–37	21·7 (7·4)	11·6 (7·1)	<0·02
Mother–daughter	7				
Mother		30–50	37·7 (8·5)		
Daughter		14–44	26·4 (10·6)	11·3 (5·5)	<0·02

age at onset in offspring) was distributed as shown in fig 2. The plot of anticipation against parental age at onset is shown in fig 3, with the regression line showing a positive intercept.

There was a highly significant correlation ($r=0·71$, $p<0·001$) for age at onset between generations I and II, which can be seen in fig 4. There was also a significant correlation for age at onset between father and offspring ($r=0·94$, $p<0·001$). No such significant correlation was found between mother and offspring and between mother and daughter, respectively. When studying the inheritance from father to son there was a highly significant correlation ($r=0·97$, $p<0·001$, $n=7$) for age at onset between generations I and II.

Maternal and paternal inheritance can be seen in tables 3 and 4. There were significant differences in total episodes per year and age at onset between father and offspring ($p<0·05$, $p<0·008$), mother and offspring ($p<0·005$, $p<0·005$), and mother and daughter ($p<0·03$, $p<0·02$). There was also a significant difference in age at onset between father and son ($p<0·02$).

To search for a possible cohort effect contributing to the difference in age at onset within parent-offspring pairs, we calculated the correlation between the year of birth difference and the onset age difference within the pairs. We found the correlation to be non-significant ($r=0·01$, $p=0·97$, $n=20$), indicating an absence of cohort effect.

Discussion

Our results support the occurrence of anticipation in families with BPAD. We found evidence for both a decrease in age at onset and an increase in severity of the disease in successive generations.

In this study we have exclusively selected families with unilinear inheritance of BPAD in order to avoid the possibility of transmission of genetic material responsible for the disease from both parents. At present it is not possible to exclude the possibility that the BPAD could be a heterogeneous group containing more than one disorder and more than one mode of inheritance. We have also excluded pedigrees containing diagnoses such as schizophrenia, schizoaffective disorder, unipolar depression, and mania for the same reasons. This gives us a fair opportunity to study anticipation as a mode of inheritance in BPAD.

In a recent study of BPAD, the disease had 8·9–13·5 years earlier age at onset and was 1·8–3·4 times more severe in generation II compared with generation I.[9] In our study the disease had about 10 years earlier age at onset and was about twice as severe in the second generation, which is in good agreement with previous results.

Generation II showed a significantly higher ($p<0·03$) frequency of episodes per year during the period without lithium treatment. A significantly higher frequency ($p<0·006$) of episodes per year was also found during treatment with lithium. Therefore, we suggest that lithium

Generation pair	No	Episodes per year			
		Ra[illegible]	Mean (SD)	Diff	p
Father–offspring	9				
Father		0·2–0·7	0·38 (0·15)		
Offspring		0·2–2·5	0·77 (0·70)	0·39 (0·71)	<0·05
Mother–offspring	11				
Mother		0·1–0·6	0·25 (0·19)		
Offspring		0·2–2·0	0·80 (0·51)	0·55 (0·41)	<0·005
Father–son	7				
Father		0·2–0·7	0·40 (0·15)		
Son		0·2–2·5	0·86 (0·78)	0·46 (0·80)	<0·08
Mother–daughter	7				
Mother		0·1–0·6	0·29 (0·19)		
Daughter		0·2–2·0	0·79 (0·59)	0·50 (0·48)	<0·03

which is also an indication that the disease is more severe in the next generation.

Interestingly, we found a highly significant correlation ($p<0·001$) between age at onset in generation I compared to II, which further supports a genetic mode of transmission in BPAD. We also found evidence of only positive anticipation (fig 3). These results suggest the occurrence of anticipation rather than ascertainment bias or some other artefact.[26]

The phenomenon of imprinting, with preferential expression of the maternally or paternally inherited allele, is known to affect gene expression and development. In this study we found that paternal or maternal transmission of genetic material influenced the age at onset and severity of the disease in the second generation equally. However, the correlation coefficients for age at onset between father and offspring was higher than that between mother and offspring.

Sporadic cases are often found in many psychiatric diseases such as schizophrenia, schizoaffective disorder, bipolar affective disorder, unipolar depression, and mania. These cases are usually called non-familial with no obvious hereditary background in the family. If the same phenomenon with the decreasing of trinucleotide repeats between generations (as recently shown for myotonic dystrophy[27]) is occurring in BPAD, followed by an expansion, this could be an explanation for the occurrence of some of the sporadic cases.

Many pedigrees show a heterogeneous pattern with different diagnoses within the same pedigree. Lately, the question has again been raised whether schizophrenia and affective disorder are genetically associated.[28] An inheritance model with a differential expansion of trinucleotide repeats in the same locus for different persons may turn out to be an important candidate to explain the observed heterogeneity.

This study was supported by the Swedish Medical Research Council (B93-21X-10412, B94-21X-03921), Joint Committee North Health Region (Sweden), Stiftelsen Bror Gardelius Society Research Fund, Swedish Society of Medical Research, and Marcus Borgström's Fund.

[illegible]okur G, Tsuang M, Crowe RR. The Iowa 500: affective disorder in relatives of manic and depressed patients. *Am J Psychiatry* 1982;139:209–12.
2 Rice J, Reich T, Andreason NC, *et al.* The familial transmission of bipolar illness. *Arch Gen Psychiatry* 1987;44: 441–7.
3 Bertelsen A, Harvald B, Hauge M. A Danish twin study of manic depressive disorders. *Br J Psychiatry* 1977;130: 330–51.
4 Mendlewicz J, Rainer JD. Adoption study supporting genetic transmission in manic-depressive illness. *Nature* 1977;268: 327–9.
5 Egeland JA, Gerhard DS, Pauls DL, *et al.* Bipolar affective disorders linked to DNA-markers on chromosome 11. *Nature* 1987;325:783–7.
6 Baron M, Risch N, Hamburger R, *et al.* Genetic linkage between X-chromosome markers and bipolar affective illness. *Nature* 1987;326:289–92.
7 Kelsoe JR, Ginns EI, Egeland JA, *et al.* Reevaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* 1989;342:238–43.
8 Mellon C. Genetic linkage studies in bipolar disorder: a review. *Psychiatr Dev* 1989;2:143–58.
9 McInnis MG, McMahon FJ, Chase GA, *et al.* Anticipation in bipolar affective disorder. *Am J Hum Genet* 1993;53: 385–90.
10 Howeler CJ, Busch HFM, Gereadts JPM, *et al.* Anticipation in myotonic dystrophy: fact or fiction? *Brain* 1989;112: 779–97.
11 Harper PS, Harley HG, Reardon W, Shaw DJ. Anticipation in myotonic dystrophy: new light on an old problem. *Am J Hum Genet* 1992;51:10–16.
12 Morel BA. *Traite des dégénérescences.* Paris: J B Baillière, 1857.
13 Mott FW. Hereditary aspects of nervous and mental diseases. *BMJ* 1910;2:1013–20.
14 Mott FW. A lecture on heredity and insanity. *Lancet* 1911; 1:1251–9.
15 Fleischer B. Uber myotonische Dystrophie mit Katarakt: eine hereditäre, familiäre Degeneration. *Arch Ophthalmol* 1918;96:91–133.
16 Sherman SL, Jacobs PA, Morton NE, *et al.* Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* 1985;69:3289–99.
17 Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971–83.
18 Snell RG, MacMillan JC, Cheadle JP, *et al.* Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genet* 1993;4: 393–7.
19 Andrew SE, Goldberg YP, Kremer B, *et al.* The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genet* 1993;4: 389–403.
20 Duyao M, Ambrose C, Myers R, *et al.* Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet* 1993;4:387–92.
21 Drugge U, Andersson R, Chizari F, *et al.* Familial amyloidotic polyneuropathy in Sweden: a pedigree analysis. *J Med Genet* 1992;30:388–92.
22 Fu YH, Kuhl DP, Pizzuti A, *et al.* Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 1991;67:1047–58.
23 Fu YH, Pizzuti A, Fenwick RG, *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 1992;255:1256–8.
24 Mahadevan M, Tsilfidis C, Sabourin L, *et al.* Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 1992;255:1253–5.
25 Farmer AE, Wessely S, Castle D, McGuffin P. Methodological issues in using a polydiagnostic approach to define psychotic illness. *Br J Psychiatry* 1992;161:824–30.
26 Ridely RM, Frith CD, Crow TJ, Conneally PM. Anticipation in Huntington's disease is inherited through the male line but may originate in the female. *J Med Genet* 1988;25:589–95.
27 Abeliovich D, Lerer I, Pashut-Lavon I, *et al.* Negative expansion of the myotonic dystrophy unstable sequence. *Am J Hum Genet* 1993;52:1175–81.
28 Baron M, Gruen RS. Schizophrenia and affective disorder: are they genetically linked? *Br J Psychiatry* 1991;159: 267–70.

EXHIBIT C

Am. J. Hum. Genet. 54:864-870, 1994

Evidence for Anticipation in Schizophrenia

Anne S. Bassett* and William G. Honer†

*Department of Psychiatry, University of Toronto, and Genetics Section, Schizophrenia Research Program, Queen Street Mental Health Centre, Toronto; and †Department of Psychiatry, University of British Columbia, Vancouver

Summary

Anticipation, or increasing severity of a disorder across successive generations, is a genetic phenomenon with an identified molecular mechanism: expansion of unstable trinucleotide repeat sequences. This study examined anticipation in familial schizophrenia. Three generations of siblines from the affected side of families selected for unilineal, autosomal dominant-like inheritance of schizophrenia were studied (n = 186). Across generations more subjects were hospitalized with psychotic illness ($P<.0001$), at progressively earlier ages ($P<.0001$), and with increasing severity of illness ($P<.0003$). The results indicate that anticipation is present in familial schizophrenia. These findings support both an active search for unstable trinucleotide repeat sequences in schizophrenia and reconsideration of the genetic model used for linkage studies in this disorder.

Introduction

The clinical observation of anticipation—i.e., inherited illness that becomes more severe across successive generations—has recently been found to have a molecular basis: expanding GC-rich trinucleotide repeat sequence mutations (Harper et al. 1992; Sutherland and Richards 1992). In fragile X syndrome (Verkerk et al. 1991), myotonic dystrophy (Fu et al. 1992), spinobulbar muscular atrophy (Brook et al. 1992), spinocerebellar atrophy type 1 (Orr et al. 1993), and Huntington disease (Huntington's Disease Collaborative Research Group 1993), increasing severity of illness, earlier age at onset, and/or increasing proportion of ill individuals in successive generations are associated with longer trinucleotide expansions. Schizophrenia is another neuropsychiatric disorder that may display this anticipation phenomenon and that therefore may have familial forms caused by an unstable trinucleotide repeat.

Schizophrenia is a severe disorder characterized by social withdrawal and psychotic symptoms, such as delusions and hallucinations. The illness has a variable age at onset, often beginning in early adulthood and resulting in lifelong disabilities in social and occupational functioning. Evidence from family, twin, and adoption studies, including those using reliable diagnostic criteria (Lowing et al. 1983; Kendler et al. 1985), strongly supports a genetic etiology for schizophrenia (Gottesman and Shields 1982). However, the mode of inheritance for schizophrenia is not readily identifiable and is proposed to involve interacting genes (Risch 1990). In families with the illness, reduced penetrance and variable expression are commonly found. Other psychotic disorders, of lesser severity, and schizotypal personality traits such as social isolation, odd communication, and extreme suspiciousness are conditions likely reflecting variable expression of genetic susceptibility to schizophrenia (Gottesman and Shields 1982; Lowing et al. 1983; Kendler et al. 1985). These factors, along with the possibility of genetic heterogeneity and the practical difficulties of studying an illness with a significant suicide rate and suspicious, socially isolated individuals, combine to make schizophrenia a challenging disorder for linkage studies (Bassett 1991). Strategies to overcome these difficulties include focusing on familial schizophrenia where inheritance is consistent with Mendelian patterns, using reliable diagnostic methods, highly polymorphic DNA markers, and lod-score methods that model the complexities of the inheritance. Linkage studies to date, using informative fami-

Received September 13, 1993; accepted for publication December 22, 1993.

Address for correspondence and reprints: Dr. Anne S. Bassett, Schizophrenia Research Program, Queen Street Mental Health Centre, 1001 Queen Street West, Toronto, Ontario M6J 1H4, Canada.

0002-9297/94/5405-0016$02.00

lies and models based on Mendelian inheritance, have yielded significant lod scores but no replicated positive results (Bassett 1991).

Dynamic modifications of classical patterns of genetic transmission, such as anticipation (Mott 1911) and genomic imprinting (differential expression of genetic material depending on parental origin of the gene and the underlying molecular mechanisms), may explain the complex genetics of schizophrenia and other major mental illnesses (McInnis et al. 1991; Flint 1992). For example, trinucleotide repeats can cause reduced penetrance and variable expression (Caskey et al. 1992; Sutherland and Richards 1992), by existing in a premutation form, by reductions in repeat size, or by somatic mutation in early embryogenesis (Lavedan et al. 1993). The unstable nature of trinucleotide repeats also provides a possible mechanism for the high mutation rates proposed for schizophrenia (Slater and Cowie 1971). If there were evidence of anticipation in schizophrenia, screening for triplet repeat mutations would become a rational option for gene localization studies, and modification of the genetic model used in linkage studies would need to be considered. The current study investigated whether anticipation was present in a familial schizophrenia sample participating in a linkage study.

Subjects and Methods

Subjects were members of eight extended nonconsanguineous families participating in a genetic linkage study of familial schizophrenia. Local psychiatrists identified prospective pedigrees segregating schizophrenia. Families were selected for large size, availability of two or more generations of adults, and apparent unilineal, autosomal dominant–like inheritance of schizophrenia and genetically related disorders. Bilineal families with evidence, from family or collateral history, of schizophrenia or other nonaffective psychotic disorders on both sides were excluded. Further details of the original ascertainment and assessment for the linkage study are described elsewhere (Bassett et al. 1993). Since families were ascertained in their entirety, proband status could be assigned to all affected subjects. Therefore no subjects were excluded from analyses.

Subjects only from the affected side of each family were taken into account, to determine sibling sets (siblines) in the index generation (IG) (n = 13 siblines), parental generation (PG) (n = 10 siblines), and grandparental generation (GG) (n = 8 siblines). The affected side was defined by (1) a parent hospitalized with psychosis (four cases), (2) an aunt/uncle hospitalized with psychosis (four cases), (3) a parent/aunt/uncle with schizotypal traits (seven cases), or (4) a sibship linking two affected nuclear families (six cases in two extended families) (see fig. 1). In two cases in the GG, the affected side could not be determined, and the smaller of the unaffected maternal or paternal siblines was arbitrarily selected. There were three instances of unknown paternity. In two of these cases, the maternal line was affected. In the third case, neither the mother nor her five siblings were affected, and this GG sibline was not included in the analysis.

Family-history information was obtained for each subject from three or more family members by using the Family History–Research Diagnostic Criteria (FH-RDC) method (Andreasen et al. 1977). Genealogical records were used to confirm dates of birth and death. Medical records were searched back to 1866 for evidence of psychiatric hospitalization. Because the subject families originated and seldom moved from a circumscribed region of Canada, and because the one psychiatric hospital available until the 1980s consistently maintained a comprehensive file-card system of recording admissions, virtually complete ascertainment of psychiatric hospitalization was assured. Records were collected for all subjects with a history of psychiatric hospitalization. Living subjects participating in the linkage study were directly interviewed by a psychiatrist (Bassett et al. 1993). Diagnostic folders containing the family history and, if present, medical records and interview data were reviewed independently by two psychiatrists (A.S.B. and W.G.H.), one of whom (W.G.H.) was blind to the pedigree structures. A consensus lifetime Research Diagnostic Criteria (RDC) diagnosis for psychotic disorders, age at first hospitalization for a psychotic illness, and presence of two or more RDC schizotypal traits were recorded. Psychotic disorders included schizophrenia (n = 25), schizoaffective disorder (n = 13; 12 mainly schizophrenic type and 1 other), unspecified functional psychosis (n = 4), mania with psychosis (n = 1), and depression with psychosis (n = 1). Schizophrenia and schizoaffective disorder were approximately equivalent in severity in the current sample (Bassett et al. 1993) and were considered together in the current study. Individuals with psychotic disorders not severe enough to require hospitalization, as well as subjects with schizotypal traits, were combined in a single schizotypal group, because of small numbers in each category.

Of the 209 subjects in the affected siblines, 23 were excluded from the analyses. One IG subject had not attained the age of 15 years, considered a minimum age

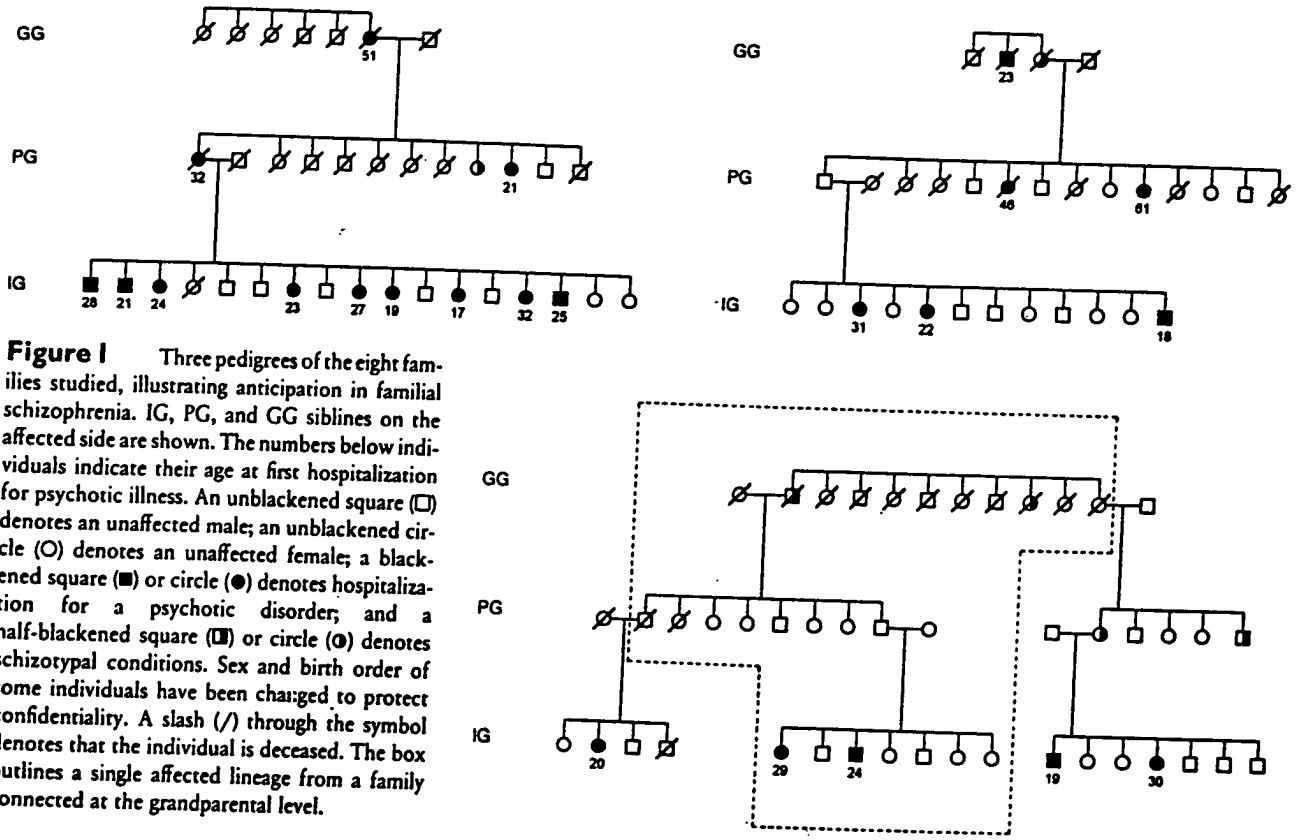


Figure 1 Three pedigrees of the eight families studied, illustrating anticipation in familial schizophrenia. IG, PG, and GG siblines on the affected side are shown. The numbers below individuals indicate their age at first hospitalization for psychotic illness. An unblackened square (□) denotes an unaffected male; an unblackened circle (○) denotes an unaffected female; a blackened square (■) or circle (●) denotes hospitalization for a psychotic disorder; and a half-blackened square (◧) or circle (◐) denotes schizotypal conditions. Sex and birth order of some individuals have been changed to protect confidentiality. A slash (/) through the symbol denotes that the individual is deceased. The box outlines a single affected lineage from a family connected at the grandparental level.

of risk for psychotic illness (Gottesman and Shields 1982). Five subjects (2 PG and 3 GG) had moved, and collateral information was insufficient to determine hospitalization status; and 17 subjects (3 IG, 5 PG, and 9 GG) died before the age of 40 years. Thirteen died in infancy or childhood, two in war, one in an accident at work, and one of unknown cause; none were suicides. Most new cases of schizophrenia may be expected before age 40 years (Gottesman and Shields 1982). Data on the remaining 186 subjects were examined for anticipation, in three ways. First, the rates of hospitalization for psychotic disorders and the rates of schizotypal conditions were compared across generations by using χ^2 analyses. Second, to assess severity of illness, subjects were assigned the following ratings: hospitalized with schizophrenia or schizoaffective disorders—3; hospitalized with other psychotic disorders—2; schizotypal—1; and unaffected—0. Means for each generation were compared using the one-way analysis of variance (ANOVA), including correction for multiple tests of significance with the Student-Newman-Keuls procedure. Third, age at first hospitalization for psychosis was assessed using the life-table method of survival analysis for 1-year intervals. Homogeneity of survival curves over the generations was examined using the Wilcoxon test. The analysis was performed assuming (1) no differential mortality between affected and unaffected and (2) hospitalization rates independent of chronological time (e.g., 1920 vs. 1970). Observations ended at the subject's current age or age at death. Covariates tested were sex and transmission patterns (maternal/paternal).

Results

Demographic characteristics of the sample and results indicating anticipation are presented in table 1. As for other illnesses demonstrating anticipation, expression of illness varied between members of a sibship (fig. 1).

Rates and Distribution of Illness

There were significantly more subjects hospitalized for psychosis across successive generations ($\chi^2 = 16.84$, $P<.0001$, 2 df). Most had schizophrenia or schizoaffective disorders (IG, $n = 30$; PG, $n = 8$; and GG, $n = 0$). Of the six subjects with less severe disorders—unspecified functional psychosis ($n = 4$), psychotic mania ($n = 1$), or psychotic depression ($n = 1$)—four were in PG or GG. Subjects with the least severe illnesses (schizotypal con-

Table 1

Characteristics of the Sample, by Generation (n = 186)

	IG	PG	GG
Total no. of subjects (females)	86 (38)	62 (24)	38 (21)
Mean size of sibline (SD)	6.61 (4.27)	6.20 (3.61)	5.43 (4.20)
Mean age of living[a] (SD)	40.61 (8.74)	65.55 (12.96)	71.80 (8.26)
Mean age at death[b] (SD)	44.25 (3.20)	57.15 (14.58)	77.35 (9.34)
Mean age at first hospitalization (SD)	26.16 (8.28)	34.00 (17.28)	41.33 (15.89)
Subjects hospitalized for psychotic illness	32 (37.21%)	9 (14.52%)	3 (7.89%)
Subjects with schizotypal conditions	8 (9.30%)	8 (12.90%)	8 (21.05%)

[a] No. of living subjects: IG, 82; PG, 42; and GG, 10.

[b] No. of dead subjects with known age at death: IG, 4; PG, 20; and GG, 20. Eight other GG subjects had ages at death that were less precisely known, e.g., "in their 70s." The mean shown did not change significantly when these subjects were included using estimated ages at death.

ditions) were twice as common in the GG as in the IG (21.05% vs. 9.30%); however, across the three generations, the result was not significant ($\chi^2 = 3.24$, $P = .20$, 2 df). When these subjects with schizotypal conditions were included with hospitalized subjects, there were still significantly more affected subjects in the youngest generation ($\chi^2 = 6.86$, $P = .032$, 2 df).

Because of the possibility of compounded error with families being connected at the parental or grandparental level, analyses were rerun with a single affected lineage from each of the seven kindreds that had data from all three generations, with the largest siblines being selected (see fig. 1). The same results were found using this subsample of 151 subjects, for both hospitalization rates ($\chi^2 = 12.30$, $P<.002$, 2 df) and hospitalization plus schizotypal rates ($\chi^2 = 9.46$, $P<.009$, 2 df). Excluding subjects who died before age 40 or moved away (n = 22) could have influenced the study's findings, since most were in the two senior generations. When results with these subjects and with penetrance estimated maximally at 100% were considered, so that one-half ($n = 11$) would have been hospitalized with psychotic disorders, the results for hospitalization rates would still have remained significant ($\chi^2 = 10.02$, $P<.01$, 2 df). Data were also reanalyzed examining the effect that increasing hospitalization rates over time would have on the results. Secular trends of hospitalization rates for specific illnesses were not available. Therefore, arbitrary increases were tested, by adding 100% more subjects ($n = 3$) to the hospitalized GG group and 50% more ($n = 4.5$) to the PG group. Under these conditions, a significant increase in the rate of psychosis requiring hospitalization would continue to be present over the generations ($\chi^2 = 7.75$, $P<.05$, 2 df).

Severity of Illness

Means for four-point severity-of-illness ratings for the three generations were as follows: IG ($n = 86$), 1.19; PG ($n = 62$), 0.58; and GG ($n = 38$), 0.37. Severity of illness significantly increased over the generations ($F = 8.39$, 2 df, $P = .0003$). Pairwise comparisons using the Student-Newman-Keuls test revealed that the significant difference was between the IG and the PG and GG. PG and GG severity means were not significantly different from each other.

Age at First Hospitalization

The survival curve (fig. 2) shows that subjects were first hospitalized for psychosis at progressively younger ages, across generations ($\chi^2 = 26.76$, $P = .0001$, 2 df). Sex of subject was not a significant covariate to generation, for age at first hospitalization (increment $\chi^2 = 0.40$, $P = .53$, 1 df). All of the GG and most (68%) of the PG subjects had achieved an age of 55 years or more, by which time they would have been virtually through the age at risk for schizophrenia (Gottesman and Shields 1982).

Imprinting

There were equal rates of maternal and paternal transmission (six and seven cases, respectively) from the PG to the IG. The mean age at first hospitalization for maternal transmission (PG to IG) was 25.00 years (SD 5.94 years), and that for paternal transmission was 27.64 years (SD 10.63 years), a nonsignificant difference ($t = .83$, $P = .41$). Grandparental to parental transmission was predominantly maternal (seven of nine cases), with one unknown and only one example of paternal transmission. Sex of transmitting parent was not a sig-

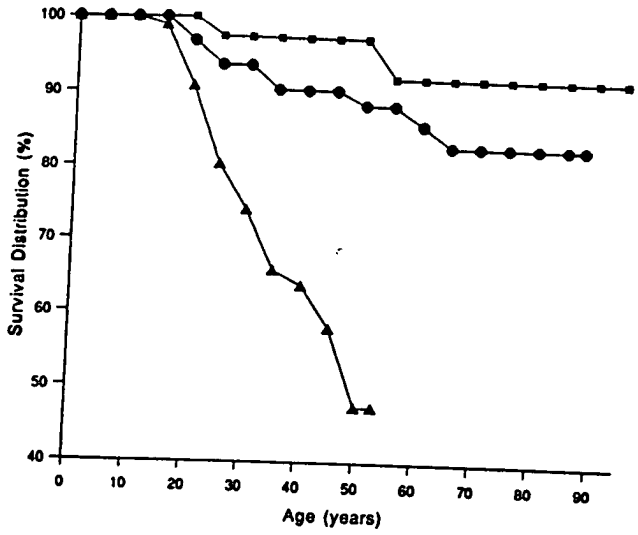


Figure 2 Survival curves for age at first hospitalization for psychotic illness, comparing IG (▲), PG (●), and GG (■). Annual percent surviving without hospitalization is plotted for every 5th year. Observation of individuals ended at their age of death or, if they were living, at their current age.

nificant covariate in the survival analysis examining age at first hospitalization (increment $\chi^2 = 0.32$, $P = .57$, 2 df).

Discussion

The results suggest that familial schizophrenia exhibits anticipation. All of the families studied showed this phenomenon, manifest as increasing rates of hospitalized psychotic illness, worsening severity of illness, and/or earlier age at onset, across successive generations (fig. 1). These findings are consistent with differences in rates of hospitalization and age-at-onset data for parent-child pairs in studies of schizophrenia over the century (Mott 1910; Kay 1963; Penrose 1971; Decina et al. 1991). In each of these studies, rates of hospitalization for psychosis were lower for antecedent generations, and age at onset for parents was significantly later than that for offspring. Investigations of ancestors and extended families also support these findings (Karlsson 1966; Odegaard 1972; Wetterberg and Farmer 1991). As well, less severe psychotic illnesses (e.g., affective disorders) are consistently more common in the generation antecedent to schizophrenic probands (Slater and Cowie 1971; Bleuler 1978). These results complement reported morbid risk of schizophrenia for parents, which is almost half that for siblings (Gottesman and Shields 1982). While alternative reasons, including selection biases such as reduced fertility in earlier-onset schizophrenia, have been proposed to explain these clinical observations, they are all consistent with the phenomenon of anticipation in schizophrenia.

In contrast, findings from the current study are only suggestive of sex-specific differences in the transmission of schizophrenia. However, the possibility of an excess maternal over paternal transmission in schizophrenia is consistent with trends found both recently by others (Sharma et al. 1993) and in studies of large data sets in the older literature (Penrose 1971; Slater and Cowie 1971). Fertility may be especially low in male patients with schizophrenia (Gottesman and Shields 1982), and in the current study this could be the reason for the high rate of maternal transmission from the GG lines to the PG lines. If this were the case, however, one would have expected predominantly maternal inheritance from the PG to the IG lines; but maternal and paternal rates were equal. Comparable transmission patterns have been found to be due to greater variation of trinucleotide repeat length after female meiosis in myotonic dystrophy (Lavedan et al. 1993). The effect may be subtle, requiring larger samples to demonstrate imprinting in schizophrenia.

There are several possible biases that can explain results that indicate anticipation (Penrose 1948). First, subjects could have died before expressing the mutation. However, all of the GG and most of the PG lived beyond age 55 years. Also, significantly different hospitalization rates across generations remained in the current study, even when half of those who had moved or died before age 40 years were assigned affected status. Second, reduced fertility of individuals with earlier onset of schizophrenia could cause preferential ascertainment of parents with later onset. This bias should be minimized in the current study, because (1) few parents were affected with psychosis and (2) siblines were large in all three generations, providing multiple opportunities for detection of affecteds, regardless of their fertility. Third, subjects in the IG could have been too young to yet express a late-onset form of psychosis, which would attenuate the age-at-onset findings. Since one-half of the subjects in the IG were over age 40 years, most were beyond the period of highest risk. Even if more new cases of psychosis subsequently arose, this would most likely occur in the IG and would only serve to strengthen the present study's findings with respect to differential rates of illness.

In contrast to major depression (Gershon et al. 1987), there is no evidence for a cohort effect in schizophrenia. However, secular trends, such as improved detec-

tion, that could, over time, lead to higher hospitalization rates and/or younger age at first hospitalization for psychosis are important to consider, since these could have influenced the principal findings of the current study. In the literature (Mott 1910; Kay 1963; Penrose 1971; Decina et al. 1991), examination of ages at first hospitalization did not reveal secular trends to younger age over the century. In the current study, arbitrary increases in hospitalization incidence assigned to the PG or GG did not change the observation of anticipation. Specific factors that may influence secular trends in hospitalization, including drug abuse and psychosocial stressors, do not appear to have played a role in the sample studied. Only three hospitalized subjects in the IG had a history of stimulant or hallucinogen use, drugs that may in some cases precipitate a psychotic illness. On the basis of direct interviews, it appeared that psychosocial stressors endured by PG and GG, such as the World Wars and Great Depression, were more severe than those faced by the IG.

Another possibility is that the family-history method tends to underestimate rates of psychiatric disturbance, particularly in relatives who are dead or less known (Andreasen et al. 1977). The consequence could be that actual rates for schizotypal traits could be higher than those found, particularly for PG and GG. However, results for severe illness requiring hospitalization would remain unchanged. Another factor that could have compounded errors in the PG and GG was the use of IG siblines connected at the parental or grandparental level. However, both the fact that results remained the same when only one ascending line from each kindred was examined and the fact that at least 13 affected siblines had resulted from eight originating GG lines support the finding of anticipation.

A limitation that is important to consider in studies of common illnesses is the potential for assortative mating to cause an apparently increased prevalence of illness in offspring. This possibility was minimized by selecting unilineal pedigrees with no evidence of schizophrenia or related disorders in the married-in person, their siblings, or parents. Although individuals marrying in could have been nonexpressing carriers of the disorder with nonexpressing close relatives, the likelihood appears small that assortative mating could account for the results in the current study. Because the families studied were selected because of their autosomal dominant–like inheritance and large sibships, the results may not be generalizable to schizophrenia in the general population, although they are consistent with observations from large population-based samples (Penrose 1971; Bleuler 1978). Also, the schizophrenia in the subject families may be a particularly severe form, which could exaggerate the findings. However, the mean age at onset for the IG is similar to others' results (Mott 1911; Gottesman and Shields 1982; Decina et al. 1991; Sharma et al. 1993). Data on specific symptom patterns suggest that the familial schizophrenia in the present families is comparable in nature and in severity to samples drawn from the general population (Bassett et al. 1993).

Despite the possible biases and limitations, the weight of evidence from both the current investigation and the literature is consistent with the finding of anticipation in familial schizophrenia. Other families with schizophrenia should be examined for anticipation and possible accompanying maternal imprinting phenomena, to confirm the current study's findings. Although published pedigrees consistently show evidence of anticipation (Karlsson 1966; Wetterberg and Farmer 1991), complete ascertainment of other large kindreds with contemporary reliable diagnostic assessments would be useful. However, the most exciting possibilities—and the confirmation of the clinical observations of the current study—lie in the search for expanding trinucleotide repeats and other DNA sequence mutations in schizophrenia. New methods becoming available to detect these mutations (Orr et al. 1993; Schalling et al. 1993) will complement and may accelerate the search for pathological genes in linkage studies. In addition, the current study has implications for the genetic model used in linkage studies. Parameters, particularly penetrance, which would vary according to generation, may need to be modified to reflect the effects of anticipation and possibly imprinting. In the light of clinical evidence for anticipation, these strategies represent real promise for deciphering the genetics of schizophrenia.

Acknowledgments

The authors are grateful to the families for their cooperation; to J. McAlduff, R.N., and A. Bury, B.A., for invaluable assistance; and to N. Risch, Ph.D., for comments on the manuscript. This work was supported in part by funding from the Scottish Rite Schizophrenia Research Program, Ian Douglas Bebensee Foundation, Ontario Mental Health Foundation, and Medical Research Council of Canada.

References

Andreasen N, Endicott J, Spitzer R, Winokur G (1977) The family history method using diagnostic criteria: reliability and validity. Arch Gen Psychiatry 34:1228–1235

Bassett AS (1991) Linkage analysis of schizophrenia: challenges and promise. Soc Biol 38:189–196

Bassett AS, Collins EJ, Nuttall S, Honer WG (1993) Positive and negative symptoms in families with schizophrenia. Schizophr Res 11:9–19

Bleuler M (1978) The schizophrenic disorders: long-term patient and family studies. Yale University Press, New Haven

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, et al (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808

Caskey CT, Pizzuti A, Fu Y, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–788

Decina P, Mukherjee S, Lucas L, Linder J, Horwath E (1991) Patterns of illness in parent-child pairs both hospitalized for either schizophrenia or a major mood disorder. Psychiatry Res 39:81–87

Flint J (1992) Implications of genomic imprinting for psychiatric genetics. Psychol Med 22:5–10

Fu Y-H, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, et al (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI (1987) Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44:314–319

Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, Cambridge

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Karlsson JL (1966) The biologic basis of schizophrenia. Charles C Thomas, Springfield, IL

Kay DWK (1963) Late paraphrenia and its bearing on the aetiology of schizophrenia. Acta Psychiatr Scand 39:159–169

Kendler KS, Gruenberg AM, Tsuang MT (1985) Psychiatric illness in first-degree relatives of schizophrenic and surgical control patients. Arch Gen Psychiatry 42:770–779

Lavedan C, Hofmann-Radvanyi H, Shelbourne P, Rabes J-P, Duros C, Savoy D, Dehaupas I, et al (1993) Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. Am J Hum Genet 52:875–883

Lowing PA, Mirsky AF, Pereira R (1983) The inheritance of schizophrenia spectrum disorders: a reanalysis of the Danish adoption study data. Am J Psychiatry 140:1167–1171

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR Jr (1991) Anticipation in bipolar affective disorder. Am J Hum Genet 53:385–390

Mott FW (1910) Hereditary aspects of nervous and mental diseases. Br Med J 2:1013–1020

——— (1911) Heredity and insanity. Lancet 3:1251–1259

Odegaard O (1972) The multifactorial theory of inheritance in predisposition to schizophrenia. In: Kaplan AR (ed) Genetic factors in schizophrenia. Charles C Thomas, Springfield, IL, pp 256–275

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet 4:221–226

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14:125–132

——— (1971) Critical survey of schizophrenia genetics. In: Howells JG (ed) Modern perspectives in world psychiatry. Oliver & Boyd, Edinburgh, pp 3–19

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus traits. Am J Hum Genet 46:222–228

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nature Genet 4:135–139

Sharma T, Brynjolfsson J, Butler R, Read T, Murphy P, Sigmundsson T, Petursson H, et al (1993) Genomic imprinting in the Icelandic/British schizophrenia study: preliminary results. Schizophr Res 9:124–125

Slater E, Cowie V (1971) The genetics of mental disorders. Oxford University Press, London

Sutherland GR, Richards RI (1992) Anticipation legitimized: unstable DNA to the rescue. Am J Hum Genet 51:7–9

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914

Wetterberg L, Farmer AE (1991) Clinical polydiagnostic studies in a large Swedish pedigree with schizophrenia. Eur Arch Psychiatry Clin Neurosci 240:188–190